Infinite Sets

Course notes from NAIL063 by doc. Kynčl

Winter semester 2025/2026

Jakub Smolík

smolikj@matfyz.cz

Date of last change: November 8, 2025

Contents

1 Review of Set Theory Basics

We will be working within Zermelo–Fraenkel (ZF) set theory; that is, Zermelo's theory Z augmented by Fraenkel's axiom schema of replacement. If we further include the axiom of choice, we obtain the much stronger theory ZFC.

As for notation, I always use the symbol $\subset$ for a proper subset (or subclass) and $\subseteq$ for a general subset (or subclass). I use $\subsetneq$ only when it is important that the two sets (or classes) are not equal. Concatenated expressions such as $a \in b \in c$ mean $a \in b \wedge b \in c$. I differentiate between the symbol for equality of two objects "$=$" and the symbol for the definition of an object "$:=$". I use the following notation for defining functions.

- $f : A \to B$ is a function with domain A and codomain B.
- $f : a \mapsto b$ denotes that f maps the set $a \in A$ to the set $b \in B$.
- $f = g \circ h$ means that $f(x) = h(g(x))$ for all suitable x.

I use the terms "function," "map," and "mapping" interchangeably.

Let us first recall some important axioms and definitions.

1.1 Sets and Classes

Definition 1.1 (Class)**.** If $\varphi(x)$ is a formula, then the expression $\{x \mid \varphi(x)\}$ is called a *class term*. It defines the "collection" of all sets x satisfying $\varphi(x)$. We call this collection the *class* determined by $\varphi(x)$.

Every set is a class, but not all classes are sets (consider the class of all sets). A class that is not a set is called a *proper class*. The major difference between sets and classes is that classes cannot be members of other classes or sets, while sets can. We can substitute class terms into logical formulas in place of free variables, but unlike sets, we cannot quantify them using $\forall$ and $\exists$. It isn't hard to show that for every formula with class terms (but without quantified class variables), there is an equivalent formula in the base language without class terms.

We will usually denote sets using small letters $a, b, c, x, y, \dots$ and classes using capital letters A, B, C, etc. The exception to this are well-ordered sets, which will often be denoted as W. Finally, the class of all sets, also called the *universal class*, is denoted by V.

1.2 Axiom Schema of Replacement

Axiom 1.2. When we take any (even a class) map F and a preimage set a, then the class of images $b = F[a]$ is also a set. Formally, if $\psi(x,y)$ is a formula, without free variables y_1, y_2 and b, then the formula

$$(\forall x)(\forall y_1,y_2)\big((\psi(x,y_1) \wedge \psi(x,y_2)) \Rightarrow y_1 = y_2\big) \Rightarrow \\ (\forall a)(\exists b) : (\forall y)\big(y \in b \Leftrightarrow (\exists x)(x \in a \wedge \psi(x,y))\big)$$

is an axiom. The formula $\psi(x,y_1)$, resp. $\psi(x,y_2)$ are created from $\psi(x,y)$ by substituting y_1, resp. y_2 for y.

The first part of this axiom says that $\psi(x,y)$ should behave like a map $y = F(x)$. In the second part, a denotes the set of preimages and b the set of corresponding images.

1.3 Axiom of Choice

The *axiom of choice*, denoted AC, is one of the most important principles in modern mathematics, with profound implications in areas such as analysis or linear algebra. It states that for any collection of nonempty sets, it is possible to choose exactly one element from each set, even if the collection is infinite. When added to Zermelo–Fraenkel set theory, it yields the much more powerful ZFC. Many theorems that seem intuitively true, such as every vector space having a basis, depend on this axiom.

However, the axiom of choice is also controversial, since it leads to counterintuitive results, such as the well-ordering principle, which claims that every set can be well-ordered, or the Banach–Tarski paradox, which provides a way to decompose a solid ball into finitely many pieces and reassemble them into two identical copies of the original.

Definition 1.3 (Choice function)**.** A *choice function* (or a *selector*) on the set x is any function $f : x \to \bigcup x$ such that

$$(\forall t \in x)(t \neq \varnothing \Rightarrow f(t) \in t).$$

We can WLOG assume that the choice function is defined on $x \setminus \{\varnothing\}$ and all $t \in \mathrm{Dom}(f)$ satisfy $f(t) \in t$.

Axiom 1.4 (Axiom of Choice)**.** Every set has a choice function.

An equivalent formulation that sounds even more intuitive is that the Cartesian product of a nonempty indexed family of nonempty sets is nonempty. Other famous conditions equivalent [1] to the axiom of choice are the well-ordering principle, Zorn's lemma and the trichotomy principle.

Showing the equivalence of the nonempty Cartesian product statement isn't hard. The other conditions are much more difficult and require the use of transfinite induction. We prove their equivalence in Sections 2.4, 2.5 and 2.6.

1.4 Natural Numbers and the Axiom of Infinity

We use Von Neumann ordinals, meaning that natural numbers are defined as

$$0 \coloneqq \varnothing,\ 1 \coloneqq \{0\},\ 2 \coloneqq \{0, 1\}, \dots, n + 1 \coloneqq \{0, 1, \dots, n\} = n \cup \{n\}.$$

Definition 1.5. The *successor function* is a mapping $S : V \to V$ defined as $v \mapsto v \cup \{v\}$. For convenience, we write $v + 1 \coloneqq S(v) = v \cup \{v\}$.

Definition 1.6. A set w is *inductive* if $0 \in w$ and for all $n \in w$ also $n + 1 \in w$.

Axiom 1.7 (Axiom of Infinity)**.** There exists an inductive set.

Definition 1.8. We define the *set of all natural numbers* as the $\subseteq$–smallest inductive set. Or, equivalently, as $\bigcap\{w \mid w \text{ is inductive}\}$. We denote it by ω.

[1] "The Axiom of Choice is obviously true, the well-ordering principle obviously false, and who can tell about Zorn's lemma?" — Jerry Bona

1.5 Well-Orderings and Initial Segments

Let us recall a very important definition: the notion of *well-ordered sets.*

Definition 1.9 (Ordering)**.** A binary relation R on the class X is a

(a) *trichotomy* if for all x,$y \in X$, either $x = y$, or $x\,R\,y$, or $y\,R\,x$,

(b) *strict order* if it is anti-reflexive, strongly anti-symmetric, and transitive on X; (note that strong anti-symmetry follows from the other two),

(c) *(partial) order* if it is reflexive, weakly anti-symmetric, and transitive on X,

(d) *total (or linear) order* if it is a trichotomous partial order on X.

If R is an ordering, then instead of $x\,R\,y$ we write $x \leq_R y$ and we call $(X, \leq_R)$ an *ordered class.* Similarly, if R is a strict ordering, then we write $x <_R y$ and we call $(X, <_R)$ a *strictly ordered class.*

Note that we can easily create a strict ordering $<_R$ from $\leq_R$ and vice versa. For this reason, we will not define properties for both strict and non-strict orderings separately, because one implicitly defines the other.

Definition 1.10. We call an element of an ordered class $(X, \leq)$ *minimal* if there is no smaller one, and we call it a *minimum* if it is smaller than all others. If a minimum exists, we denote it by $\min_{\leq}(X)$. The *supremum* of a subset $Y \subseteq X$ is the minimum of all its upper bounds. If it exists, we denote it by $\sup_{\leq}(Y)$

Observation 1.11. *Every minimum is minimal. Furthermore, if R is a total order, then there is at most one minimal element, and if it exists, then it is also the minimum. There is always at most one minimum.*

Definition 1.12 (Well-ordering)**.** An ordered class $(A, \leq_R)$ is *well-ordered* if every non-empty subset of A has a minimum. Notice that every well-ordered class is totally ordered since we can take any two elements, and one of them has to be the minimum and is therefore smaller.

Observation 1.13. *The well-ordering property is hereditary. That is, if X is well-ordered by $\leq_R$, then every $Y \subseteq X$ is also well-ordered by $\leq_R$.*

Observation 1.14. *Well-ordered sets have no infinite strictly decreasing sequences — the set of elements of such a sequence would have no minimum.*

Remark. The reverse implication can't be proven in ZF; it requires a weak form of the axiom of choice, known as the axiom of dependent choice.

Definition 1.15 (Lower part and subset)**.** Let $(A, <_R)$ be a (strictly) ordered class. A subclass $X \subseteq A$ is a *lower part* of A if

$$(\forall x \in X)(\forall a \in A)(a <_R x \Rightarrow a \in X).$$

Additionally, if X is a set, we call it a *lower subset* of A, and if $X \neq A$, then we call it a *proper lower part*, or *proper lower subset* of A.

Lemma 1.16. *Let $(W, <_R)$ be a (strictly) well-ordered set, and suppose that X is a proper lower subset of W. Then there exists a unique $x \in W$ such that X is equal to the set $\{y \in W \mid y <_R x\}$. We denote this set as $(\leftarrow, x)$.*

Proof. We define x as the minimum of $W \setminus X$. Then every $y <_R x$ belongs to X, so $(\leftarrow, x) \subseteq X$. We also want the opposite inclusion. For contradiction, suppose there is a $y \in X$ such that $y \notin (\leftarrow, x)$. If $y \not< x$, then necessarily $x \leq_R y$. But this means that $x \in X$ because X is a lower subset and $y \in X$. But this is a contradiction since $x \notin X$. □

Definition 1.17 (Initial segment)**.** If $(W, <_R)$ is a (strictly) well-ordered set, then we call its proper lower subsets *initial segments* instead. We denote the unique initial segment of W determined by $x \in W$ as

$$(\leftarrow, x) \coloneqq \{y \in W \mid y <_R x\}.$$

It contains all the elements of W from the minimum of W until x, but not x itself.

Observation 1.18. *Note that $x <_R y \iff (\leftarrow, x) \subset (\leftarrow, y)$.*

2 Ordinal Numbers

Informally, *ordinal numbers* are a way to generalize natural numbers. We will first do a quick recap of the basics of ordinal numbers and then prove a theorem that deeply links ordinals and well-ordered sets.

2.1 Ordinals as a Generalization of Naturals

Definition 2.1. A class X is called *transitive* if for all $x \in X$ we have $x \subseteq X$. Or equivalently, if for every x, y such that $y \in x \in X$ we have $y \in X$.

Theorem 2.2. *Every natural number and the set of all natural numbers ω is transitive and (strictly) well-ordered by the membership relation $\in$.*

From now on, we will denote the (strictly) well-ordered set $(\omega, \in)$ as $(\omega, <)$ instead and write $n < m$ instead of $n \in m$ when talking about natural numbers.

Definition 2.3 (Ordinal numbers)**.** A set α is an *ordinal number* if it is transitive and (strictly) well-ordered by the membership relation $\in$. If α is infinite, we say that it is a *transfinite ordinal.* We denote the *class of all ordinal numbers* by On.

Theorem 2.4. *Finite ordinals are exactly the natural numbers, and ω is the smallest transfinite ordinal.*

Theorem 2.5. *The class* On *itself is transitive and (strictly) well-ordered by $\in$. This implies that it is not a set; otherwise,* On $\in$ On*. Furthermore, any proper class X that is transitive and well-ordered by $\in$ is identical to* On.

As for notation, we will use the symbols $\alpha, \beta, \gamma, \ldots$ to denote ordinals and the symbol $<$ to compare them. That is, we write $\beta < \alpha$ instead of $\beta \in \alpha$.

Observation 2.6. *If $\beta < \alpha$, then $\beta \subset \alpha$ and β is an initial segment of α. Additionally, $\alpha = (\leftarrow, \alpha)$.*

Definition 2.7. If α is an ordinal, then we call all $\beta < \alpha$ the *predecessors* of α. The *successor* of α is the ordinal $\alpha^+ := \alpha \cup \{\alpha\}$. We say that α is the *direct predecessor* of α^+.

Remark. It is easy to show that α^+ is the smallest ordinal larger than α.

Definition 2.8. We say that an ordinal number α is an

(a) *isolated* (or *successor*) ordinal if $\alpha = 0$ or α has a direct predecessor,

(b) a *limit* ordinal otherwise.

Example. Every $n \in \omega$ is isolated, ω is limit, and ω^+ is isolated again.

2.2 Ordinals as Types of Well-Ordered Sets

The definition of ordinals presented above was formalized by John von Neumann in 1923. This elegant approach, however, came decades after Georg Cantor first introduced ordinals (around 1885) as *order types of well-ordered sets.* Cantor's intuition was that ordinals serve as labels for well-ordered sets: the smallest element is labeled 0, the next 1, and so on. The *order type* of the set is then the first label we did not have to use; it represents the "shape" of the ordering.

Consider, for example, a set ordered as

$$a_0 < a_1 < a_2 < \overbrace{\cdots}^{\infty} < b.$$

Here, there are countably infinitely many elements a_i, followed by one additional element b. If we label the elements from left to right, all the natural numbers are used for the a_i's, leaving no finite label for b. This is precisely why we need transfinite ordinals: we assign the label ω to b. Hence, the order type of this ordering is $\omega^+ = \omega + 1$.

It is important to realize that different orderings of the same sets can have different order types. This means that the ordinal numbers do not count the number of objects in the set; they only label them.

Lemma 2.9. *Every proper lower part of* $(\mathrm{On}, <)$ *is an ordinal number.*

Proof. Let X be a proper lower part of On. Then

(i) X is transitive. Suppose $\alpha \in \beta \in X$, that is $\alpha < \beta \in X$. Because X is a lower part, we have $\alpha \in X$.

(ii) X is well-ordered by $\in$ because On is well-ordered by $\in$ and $X \subseteq \mathrm{On}$.

We also need to argue that X is a set. If it were a proper class, then by Theorem 2.5 it would be the entire On, but $X \subsetneq \mathrm{On}$. □

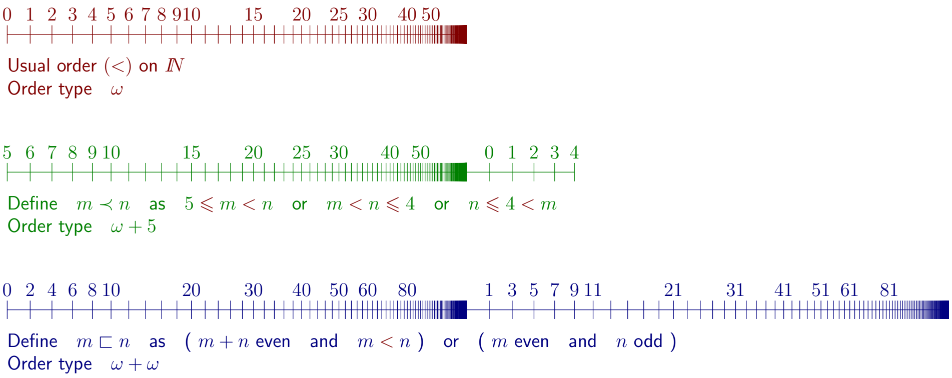

Figure 1: Different orderings of the same set can have different types; [5].

Definition 2.10 (Isomorphism)**.** Let $(A, \leq_R)$ and $(B, \leq_S)$ be ordered classes. A bijection $F : A \to B$ is an *order-isomorphism* of $(A, \leq_R)$ and $(B, \leq_S)$ if

$$(\forall x,y \in A)\big(x \leq_R y \iff F(x) \leq_S F(y)\big).$$

Because we will not be dealing with other types of isomorphisms, we will usually simply say *isomorphism* instead of order-isomorphism.

Theorem 2.11 (About comparing well-orderings)**.** *If $(W_1, \leq_1)$ and $(W_2, \leq_2)$ are well-ordered sets, then exactly one of the following holds:*

(a) *either W_1 and W_2 are isomorphic, or*

(b) *W_1 is isomorphic to an initial segment of W_2, or*

(c) *W_2 is isomorphic to an initial segment of W_1.*

In each case, the isomorphism is unique.

Corollary 2.12. *No two distinct ordinal numbers can be isomorphic.*

Proof. Suppose $\alpha < \beta$, that is $\alpha \in \beta$ and $\alpha \subset \beta$. Clearly α is an initial segment of β. This means that we are in case (b) of the previous theorem. □

Theorem 2.13 (About the type of well-ordering)**.** *Every well-ordered set is isomorphic to a unique ordinal number, which is called the* type *of the ordering.*

Proof (following [15, Thm. 3.1, Chap. 6]*).* Let $(W, <_R)$ be a well-ordered set. We want to show that there is a unique ordinal $(\alpha, <)$ isomorphic to it. Define X as the set of all $x \in W$ for which $(\leftarrow, x)$ is isomorphic to an ordinal. As no two distinct ordinals are isomorphic, this ordinal is uniquely determined, and we denote it α_x; we denote the isomorphism as $i_x : (\leftarrow, x) \to \alpha_x$.

Suppose that there exists a set S such that $S = \{\alpha_x \,|\, x \in X\} \subseteq \mathrm{On}$. Because we assume that S is a set, then $S \subsetneq \mathrm{On}$. We claim that S is a proper lower part of $(\mathrm{On}, <)$, and thus, by Lemma 2.9, it is an ordinal; let's call it α. Indeed, suppose $\beta < \alpha_x \in S$, we want $\beta \in S$. Note that β is an initial segment of α_x. This implies that $i_x^{-1}[\beta]$ is an initial segment of W. Because W is well-ordered, $i_x^{-1}[\beta]$ is equal to $(\leftarrow, b)$ for some $b \in W$ (using Lemma 1.16). So $\beta = \alpha_b \in S$ by the definition of S. More precisely, $i_x \upharpoonright (\leftarrow, b)$ is an isomorphism between $(\leftarrow, b)$ and β. We will argue why we can make the assumption that S is a set later.

A similar argument shows that X is a lower subset of W. To show this, suppose $x \in X$ and take $y \in a$ such that $y <_R x$. We want $y \in X$. We have $y <_R x$; therefore, $(\leftarrow, y)$ is an initial segment of $(\leftarrow, x)$. Because isomorphisms conserve all ordering properties, $i_x \upharpoonright (\leftarrow, y)$ is an isomorphism between $(\leftarrow, y)$ and an initial segment of α_x. By Lemma 2.9, this is an ordinal; by our previous notation, α_y. Therefore $y \in X$.

We conclude that either $X = W$ or $X = (\leftarrow, c) \subset W$ for some $c \in W$ (using Lemma 1.16). We now define a function $f : X \to S = \alpha$ by $f : x \mapsto \alpha_x$. From the definition of S and the fact that $x < y$ implies $(\leftarrow, x) \subset (\leftarrow, y)$ and therefore $\alpha_x < \alpha_y$, it is obvious that f is an isomorphism of $(X, <_R)$ and $(\alpha, <)$. If

- $X = (\leftarrow, c)$, then by the definition of the set X, $c \in X$ because $(\leftarrow, c)$ is isomorphic to an ordinal $\alpha_c = \alpha$. But this is a contradiction because $c \notin (\leftarrow, c) = X$.
- Therefore, $X = W$ and α is the sought-after ordinal isomorphic to $(W, <_R)$.

The uniqueness of α follows from the simple observation that if W were isomorphic to two distinct α_1 and α_2, then by the transitivity of isomorphisms, the ordinals α_1 and α_2 would be isomorphic, which is impossible by Corollary 2.12.

This would complete the proof if we were justified to make the assumption that the class S is a set and therefore an ordinal. In fact, we have to use the axiom of replacement to guarantee it. If we assume this axiom, then S is a set because it is the image of the set X under the map f. □

Exercise 1. Is there a well-ordered proper class not isomorphic to (On, $<$)?

Hint. Try to modify On to contradict the property described in Lemma 2.9. If you run out of ideas, consult Section 3.2.

2.3 Transfinite Induction and Recursion

In mathematics, we often use induction on the natural numbers to prove statements, and we can use recursion, such as $f(0) = 1$ and $f(n) = n \cdot f(n-1)$, to define functions. We will now show how to generalize this to all ordinals.

Theorem 2.14 (Transfinite Induction Principle)**.** *Let $A \subseteq \mathrm{On}$ be a class such that for all ordinals $\alpha \in \mathrm{On}$, we have $\alpha \subseteq A \Rightarrow \alpha \in A$, or in other words*

$$(\forall \beta < \alpha)(\beta \in A) \implies (\alpha \in A). \tag{2.1}$$

Then $A = \mathrm{On}$.

Equivalently, assume that $\varphi(x)$ is a property, and for all ordinals α:

If $\varphi(\beta)$ holds for all $\beta < \alpha$, then $\varphi(\alpha)$.

Then $\varphi(\alpha)$ holds for all ordinals α.

Proof. Suppose that $\gamma \in \mathrm{On} \setminus A$ and let $S = \{\alpha \leq \gamma \,|\, \alpha \notin A\}$. Because ordinals are well-ordered, the set S has a minimum element α. Since every $\beta < \alpha$ is in A, it follows by (2.1) that $\alpha \in A$, which is a contradiction.

The equivalence can be easily seen by taking the class $A = \{x \,|\, \varphi(x)\}$ or the property $\varphi(x) = x \in A$. □

We can also formulate the principle separately for isolated and limit ordinals, which allows us to use the transfinite induction principle in a form closer to the usual formulation of the induction principle for the naturals.

Theorem 2.15 (Transfinite Induction Principle II)**.** *Let* $A \subseteq \mathrm{On}$ *be a class satisfying*

(i) $0 \in A$,

(ii) $\alpha \in A \Rightarrow \alpha^+ \in A$, *. . . this is just induction on* ω

(iii) *if* α *is a limit ordinal and* $(\forall \beta < \alpha)(\beta \in A)$, *then* $\alpha \in A$.

Then $A = \mathrm{On}$. *Note that we can again easily reformulate this in terms of a property* $\varphi(x)$.

Proof. We need to show that these three assumptions imply (2.1). So let α be an ordinal such that $\beta \in A$ for all $\beta < \alpha$. If $\alpha = 0$, then $\alpha \in A$ by (i). If $\alpha \neq 0$ is isolated, that is if there is a $\beta < \alpha$ such that $\alpha = \beta^+$, we know that $\beta \in A$, so $\alpha \in A$ by (ii). If α is a limit ordinal, we have $\alpha \in A$ by (iii). □

We can use transfinite induction to prove properties of certain infinite structures. On the other hand, transfinite recursion—the technique described in the following theorem—allows us to construct various infinitely complex structures and define functions in a recurrent fashion.

Theorem 2.16 (About construction by transfinite recursion)**.** *If* $G : V \to V$ *is a class map, then there is a unique class map* $F : \mathrm{On} \to V$ *satisfying*

$$F(\alpha) = G(F \restriction \alpha). \tag{2.2}$$

So we define the image of the next ordinal using its predecessors and their images.

Remark. This should seem a bit suspicious because it looks like we are saying that for every class G, there exists a class F for which something holds. But we cannot quantify classes. Well, we can replace the quantification of G with a theorem *schema*, one for each G. And we aren't really quantifying F because the following proof explicitly constructs it.

Remark. The theorem can be equivalently formulated using different recurrences; for example, as

- $F(\alpha) = G\big(F[\alpha]\big) = G\big(\{F(\beta) \,|\, \beta < \alpha\}\big)$,
- $G : \mathrm{On} \times V \to V$ and $F(\alpha) = G(\alpha, F \restriction \alpha)$,
- $F(\alpha)$ is $G_1(F(\beta))$ if $\alpha = \beta^+$ is isolated, and $G_2(F[\alpha])$ if α is limit.

Additionally, these transfinite recursion statements are equivalent to the axiom of replacement, as shown in [14].

Proof. We define A as the class of "set approximations" of F. That is set mappings f, the domain of which is some ordinal number β, and that for all $\alpha < \beta$, we have $f(\alpha) = G(f \restriction \alpha)$. Now we define F as $F \coloneqq \bigcup A$. Clearly $F \subseteq \mathrm{On} \times V$. We will show that $F : \mathrm{On} \to V$ is the unique mapping satisfying (2.2).

First, we show that the approximations of F agree. Let f,$f' \in A$ and $\alpha \in \mathrm{Dom}(f) \cap \mathrm{Dom}(f')$. We claim that $f(\alpha) = f'(\alpha)$. Note that $\mathrm{Dom}(f) \cap \mathrm{Dom}(f')$ is an ordinal δ. For contradiction, suppose that $\alpha \in \delta$ is the smallest ordinal for which $f(\alpha) \neq f'(\alpha)$. Then $f \restriction \alpha = f' \restriction \alpha$ so $f(\alpha) = G(f \restriction \alpha) = G(f' \restriction \alpha) = f'(\alpha)$, a contradiction.

Second, we verify that F satisfies (2.2); that is, for all $\alpha \in \mathrm{Dom}(F)$, we have $F(\alpha) = G(F \restriction \alpha)$. So let $\alpha \in \mathrm{Dom}(F)$. It is there due to some $f \in A$ satisfying $\alpha \in \mathrm{Dom}(f)$ and $f(\alpha) = G(f \restriction \alpha)$. Also, $F(\alpha) = f(\alpha)$ and $F \restriction \alpha = f \restriction \alpha$. Therefore, by combining these equalities $F(\alpha) = G(F \restriction \alpha)$.

Next, we show that $\mathrm{Dom}(F) = \mathrm{On}$. First, we prove that $\mathrm{Dom}(F)$ is a lower part of On. Suppose $\alpha \in \mathrm{Dom}(F)$; then it is there thanks to some $f \in A$ with domain $\delta > \alpha$. If $\beta < \alpha$, then also $\beta \in \delta$, and thus $\beta \in \mathrm{Dom}(F)$.

According to Lemma 2.9, either $\mathrm{Dom}(F) = \mathrm{On}$, which we want, or $\mathrm{Dom}(F) = \gamma \in \mathrm{On}$. Suppose, for contradiction, that $\mathrm{Dom}(F) = \gamma$. Then F is a set because $\mathrm{Dom}(F)$ is a set, $\mathrm{Rng}(F)$ is a set using the axiom of replacement, and $F \subseteq \mathrm{Dom}(f) \times \mathrm{Rng}(f)$. This implies that $F \in A$ because its domain is an ordinal, and we have verified that it satisfies the recursive definition property.

Now that $F \in A$, we define a slightly "longer" function $F_1 \coloneqq F \cup \{(\gamma, G(F))\}$; note that $F = F_1 \restriction \gamma$. Notice that $F_1 \in A$ because $\mathrm{Dom}(F_1) = \gamma^+$ is an ordinal, and we defined it to satisfy the recursive definition property. Because $F = \bigcup A$, this implies $F_1 \subseteq F$, but then $\gamma \in \mathrm{Dom}(F_1) \subseteq \mathrm{Dom}(F) = \gamma$, which is a contradiction. We conclude that $\mathrm{Dom}(F) = \mathrm{On}$.

Finally, we prove the uniqueness of F. For contradiction, suppose that there is another mapping $F' \neq F$ satisfying this theorem. Because $(\mathrm{On}, <)$ is well-ordered, we can take the smallest ordinal α where $F(\alpha) \neq F'(\alpha)$. Therefore $F \restriction \alpha = F' \restriction \alpha$ and so $F(\alpha) = G(F \restriction \alpha) = G(F' \restriction \alpha) = F'(\alpha)$, which is a contradiction. $\square$

Exercise 2. Prove by induction on ω that every infinite well-ordered set A, such that each initial segment $(\leftarrow, a)$ is finite, is isomorphic to $(\omega, <)$.

Hint. Since each $(\leftarrow, a)$ is finite, there is a unique $n_a \in \omega$ with the same cardinality. The isomorphism we are looking for is $f : A \to \omega$ defined by $f : a \mapsto n_a$.

Exercise 3. Prove by transfinite recursion that every well-ordered proper class W, such that each proper lower part $(\leftarrow, a)$ is a set, is isomorphic to $(\mathrm{On}, <)$.

Hint. Use transfinite recursion to define an isomorphism $F : \mathrm{On} \to W$ using $G(x) = \min(W \setminus x)$ as $F(0) = \min(W)$ and $F(\alpha) = G(F[\alpha])$.

We will use transfinite recursion to prove the equivalence of the well-ordering theorem and Zorn's lemma to the axiom of choice. But transfinite recursion can also be used to prove some wildly sounding geometrical claims, such as

- $\mathbb{R}^3$ is a union of pair-wise disjoint unit circles, or that
- there is a set in $\mathbb{R}^2$ that intersects every line in exactly two points.

2.4 The Well-Ordering Principle

The *well-ordering principle*—the statement that every set can be well-ordered—was a foundational belief of Georg Cantor, but he was unable to provide a proof for it. This challenge was famously solved by Ernst Zermelo in 1904. Zermelo was the first person to explicitly state the axiom of choice, which he identified as the principle Cantor (and many others) had been implicitly using in many proofs. He then demonstrated that AC and the well-ordering principle are equivalent, which is why the principle is now often called the "Well-Ordering Theorem" or "Zermelo's Theorem." Veritasium has a great video [24] on this topic.

Principle 2.17 (Well-Ordering Principle)**.** Every set can be well-ordered.

Theorem 2.18. *The well-ordering principle is equivalent to the axiom of choice.*

Proof. $\mathsf{WO} \Rightarrow \mathsf{AC}$. Let $A \neq \varnothing$ be a set, without loss of generality $\varnothing \notin A$. We want to construct a selector $f : A \to \bigcup A$ such that for all $a \in A$ we have $f(a) \in a$. The well-ordering principle guarantees a well-ordering $\leq$ on $\bigcup A$, and because every a is a nonempty subset of $\bigcup A$, it has a least element with respect to $\leq$. We chose this minimum as $f(a)$.

$\mathsf{AC} \Rightarrow \mathsf{WO}$. Let $A \neq \varnothing$ be a set. We will use transfinite recursion to label the elements of A by ordinal numbers and then use the well-order of the ordinals to define a well-order on A. Let $g : \mathcal{P}(A) \to A$ be a selector on $\mathcal{P}(A)$, assigning to each nonempty $B \subseteq A$ an element $b \in B$. We will want to use transfinite recursion based on g, so we should extend it to be a class map $G : V \to V$, for example, by defining it to be equal to $\varnothing$ when g is not defined.

We can now use transfinite recursion to define the function $F : \mathrm{On} \to A \cup \{\varnothing\}$ as $F(0) = G(A)$ and $F(\alpha) = G(A \setminus F[\alpha])$. This function assigns to each ordinal a unique element from A until they "run out" (when $F[\alpha] = A$), and then it assigns $\varnothing$ to all larger ordinals.

Define W as the class of all ordinals α for which $F[\alpha] \subsetneq A$. Denote the restriction of F to W as $F_W : W \to A$. Plan: first, we show that W itself is an ordinal. From this, it will follow that F_W is a bijection between W and A, allowing us to denote the unique ordinal mapped to $a \in A$ as α_a. Once this is established, we define a well-ordering R of A as

$$a <_R b \iff \alpha_a < \alpha_b.$$

This is a well-ordering since $(A, <_R)$ is order-isomorphic to $(W, <)$, which is well-ordered (as W is an ordinal).

Firstly, we claim that W is a set. Indeed, because F_W is injective, it has an inverse F_W^{-1} that maps the set $\mathrm{Rng}(F_W) \subseteq A$ onto W, which is therefore, using the axiom of replacement, a set. Now we claim that W is a lower subset of On, and so it is an ordinal (by Lemma 2.9). Suppose $\alpha \in W$, that is $F[\alpha] \subsetneq A$, and let $\beta < \alpha$. Then $\beta \subseteq \alpha$ and $F[\beta] \subseteq F[\alpha]$, so $\beta \in W$.

To complete the proof, we must show that $F_W : W \to A$ is a bijection. It is clearly injective. To show that it is surjective, suppose for contradiction that there exists some $b \in A \setminus F_W[W]$. Because W is an ordinal number γ, it satisfies the definition of W (thanks to b) and thus $W = \gamma \in W$, which is a contradiction. $\square$

2.5 Zorn's Lemma

Zorn's lemma is perhaps the most useful application of the axiom of choice outside of set theory. It is also known as the maximality principle, a name that dates back to the German mathematician Felix Hausdorff, who proved an earlier and equivalent version of the theorem in 1914 (see [27] for details). The formulation known today as Zorn's lemma was introduced in 1935 by the German mathematician Max Zorn. However, it had already been independently proven in 1922 by the Polish mathematician Kazimierz Kuratowski, whom you might know for Kuratowski's theorem—a forbidden-graph characterization of planar graphs.

Definition 2.19 (Chain)**.** Let $(a, \leq_R)$ be an ordered set. We call the subset $b \subseteq a$ a *chain* if b is totally ordered by $\leq_R$.

Principle 2.20 (Zorn's Lemma)**.** Every (partially) ordered set containing upper bounds for every chain necessarily contains at least one maximal element.

There is also a parameterized version of this statement.

Principle 2.21 (Parametrized Zorn's Lemma)**.** Let A be a (partially) ordered set containing upper bounds for every chain. Then for every $a \in A$, there is a maximal element $b \in A$ such that $a \leq b$.

We can obtain the parameterized version from the unparameterized one by restricting ourselves to the elements above or equal to a. The other direction is obvious.

Remark. Zorn's lemma can be made slightly stronger by assuming that only well-ordered chains have upper bounds. The proof remains virtually unchanged.

Theorem 2.22. *The axiom of choice implies Zorn's lemma.*

Proof. Let $(A, <_R)$ be an ordered set containing upper bounds for each chain and for contradiction suppose that there is no maximal element. Note that this implies that every chain, in fact, has a *strict* upper bound. If a chain C had no strict upper bound, then the non-strict upper bound $b \in C$ would be a maximal element. We denote the set of strict upper bounds of C as $C^{>}$.

We take $f : \mathcal{P}(A) \to A$, a selector on $\mathcal{P}(A)$, and define a function g from the set of all chains in A as $g(C) := f(C^{>})$. So g maps a chain to one of its strict upper bounds. Now pick an arbitrary $a \in A$ and define the mapping $H : \mathrm{On} \to A$ by transfinite recursion as $H(0) = a$ and $H(\alpha^+) = g(\{H(\alpha)\})$ for successor ordinals, and as $H(\delta) = g(H[\delta])$ for limit ordinals. We start with a and get larger and larger elements of A using successor ordinals, each time taking a strict upper bound of a single element chain. If an ordinal δ is limit, we notice that $H[\delta]$ is a chain (all the smaller elements that we picked previously are strict upper bounds of each other and are therefore comparable), and $H(\delta)$ is a strict upper bound of this chain.

Note that if we want to be rigorous about the construction by transfinite recursion, we should define g on the entire V. But we can do this in any way, for example, by defining $G(x)$ as $\varnothing$ if x is not a chain of A, and $g(x)$ otherwise.

Finally, observe that $H : \mathrm{On} \to A$ is an increasing function (each value is a strictly larger upper bound than the previous one) and that it is injective.

Thus, we obtain an injection from the proper class On into the set A, which is impossible. Indeed, taking the inverse mapping and applying the axiom of replacement would imply that On itself is a set, which is a contradiction. $\square$

Theorem 2.23. *Zorn's lemma implies the well-ordering principle.*

Proof. Let X be any set. We will find a well-ordering of it by considering all of its possible well-ordered subsets, picking the maximal one using Zorn's lemma, and showing that it orders the entire X. Consider the set: [2]

$$\mathcal{W} \coloneqq \{(A, <_R) \mid <_R \text{ is a well-order on } A \subseteq X\},$$

and define a partial order $\prec_{\mathcal{W}}$ on it by $(A, <_R) \prec_{\mathcal{W}} (B, <_S)$ if B end-extends A. That is, if $A \subset B$, and $<_R$ is the restriction of $<_S$ to A, and A is an initial segment of B. We will apply Zorn's lemma to $\mathcal{W}$.

First, we need to show that chains have upper bounds. Let $\mathcal{C} \subseteq \mathcal{W}$ be a chain. Define the set

$$M \coloneqq \bigcup \{A \mid (A, <_R) \in \mathcal{C}\} \subseteq X,$$

and for $x, y \in M$ put $x <_M y$ if there exists some $(A, <_R) \in \mathcal{C}$ such that $x, y \in A$ and $x <_R y$. Because $\mathcal{C}$ is a chain, this is well-defined: if x and y belong to two distinct orderings in $\mathcal{C}$, then one extends the other and hence they agree.

We claim that $(M, <_M)$ is well-ordered. Let $S \subseteq M$ be nonempty and pick some $s \in S$. Then $s \in A_s$ for some $(A_s, <_R) \in \mathcal{C}$. Note that $A_s \cap S$ is nonempty, and because A_s is well-ordered, there exists a minimum $m = \min_{<_R}(A_s \cap S)$. Notice that also $m = \min_{<_M}(S)$. Indeed, if there were a $t \in S \setminus A_s$ such that $t <_M m$, then it would be in S due to some $A_t \in \mathcal{C}$ containing t. Since both A_s and A_t are in the chain, either

(a) $A_t \subseteq A_s$, which is impossible since then $t \in A_s$, contradicting the minimality of m in $A_s \cap S$, or

(b) $A_s \subset A_t$, meaning that A_s is an initial segment of A_t, and therefore $m \in A_s$ is smaller than $t \in A_t \setminus A_s$, which contradicts the assumption that $t <_M m$.

Therefore $(M, <_M)$ is well-ordered and thus an upper bound of $\mathcal{C}$ in $\mathcal{W}$.

Because all chains are bounded, by Zorn's lemma, $\mathcal{W}$ has a maximal element $(W, <_W)$. We claim that $W = X$ and so it is the sought-after well-ordering of X. For contradiction, suppose there exists some $x \in X \setminus W$ and extend the ordering $<_W$ to $W' \coloneqq W \cup \{x\}$ by making each $y \in W$ smaller than x. Notice that this slightly "longer" order is a well-ordering of W' and therefore is in $\mathcal{W}$. Moreover, it end-extends $(W, <_W)$ which hence is not maximal in $(\mathcal{W}, \prec_{\mathcal{W}})$. We have arrived at a contradiction and can conclude that $W = X$. $\square$

Exercise 4. Would the proof still have worked if instead of end-extensions, we had simply used general extensions? Meaning that the smaller ordering doesn't need to be an initial segment of the larger one.

Hint. By defining the end-extension ordering, we have ensured that chains have a similar structure to chains of ordinals (larger ordinals end-extend the smaller ones). Thus, when proving that M is well-ordered, we could have used a similar strategy as when proving that the ordinals are well-ordered.

[2] Why is this a set?

To demonstrate an application of Zorn's lemma, consider the following question. Does every connected graph have a spanning tree? Finding one in a finite graph is easy: simply remove the edges of cycles until there are no cycles left. But this process may not terminate for infinite graphs.

Proposition 2.24. *Every connected graph has a spanning tree.*

Sketch of proof. The set of all sub-graphs that are trees is partially ordered by inclusion, and the union of a chain is its upper bound. Zorn's lemma states that a maximal tree must exist, which is a spanning tree since the graph is connected. □

Remark. In general, suppose that we have a structure represented by a set X (a graph) with substructures $A \subseteq X$ (subgraphs that are trees), and we want to show that there is a maximal substructure. Then we simply need to check that the union of a chain of substructures is itself a substructure.

2.6 The Trichotomy Principle

Definition 2.25. For sets x and y we define the relations

(a) $x \approx y$, if there exists a bijection $x \to y$,

(b) $x \preceq y$, if there exists a injection $x \to y$,

(c) $x \prec y$, if $x \preceq y$ and $x \not\approx y$.

Theorem 2.26 (Cantor, Berstein)**.** $x \approx y \iff (x \preceq y \wedge y \approx x)$.

Principle 2.27 (Trichotomy principle)**.** The relation $\preceq$ is is trichotomous on V. That is, for any sets x and y either $x \preceq y$, or $y \preceq x$.

Theorem 2.28. *Zorn's lemma implies the trichotomy principle.*

Proof. Let x, y be arbitrary sets; we want to find an injection $x \to y$ or $y \to x$. Consider the set[3]

$$\mathcal{F} = \{f \mid f \text{ is an injection, } \mathrm{Dom}(f) \subseteq x \text{ and } \mathrm{Rng}(f) \subseteq y\}.$$

Notice that the ordered set $(\mathcal{F}, \subseteq)$ satisfies the conditions of Zorn's lemma since the union of a chain of injections is again an injection. Let g be a maximal element of $\mathcal{F}$. If both $x \setminus \mathrm{Dom}(g)$ and $y \setminus \mathrm{Rng}(g)$ were non-empty, then it would be possible to extend g by an extra pair, contradicting its maximality. Hence either $\mathrm{Dom}(g) = x$ and then $x \preceq y$, or $\mathrm{Rng}(g) = y$ and then $y \preceq x$. Here, we used the fact that the inverse of an injection is also an injection. □

Later, in Theorem 3.42, we will show that the trichotomy principle implies the well-ordering principle and thus also the axiom of choice.

Theorem 2.29. *We conclude that the following statements are equivalent in* ZF*:*

- *the axiom of choice,*
- *the well-ordering principle,*
- *Zorn's lemma,*
- *the trichotomy principle.*

[3] Why is this a set?

3 Operations on Ordinals

3.1 Ordinal Functions

Definition 3.1 (Ordinal function)**.** We say that a mapping F is an *ordinal function* if its domain is a lower part of On, that is $\mathrm{Dom}(F) \in \mathrm{On}$ or $\mathrm{Dom}(F) = \mathrm{On}$, and $\mathrm{Rng}(F) \subseteq \mathrm{On}$. We say that F is

(a) *increasing* if for all $\beta \in \mathrm{Dom}(F)$ and $\alpha < \beta$ we have $F(\alpha) < F(\beta)$,

(b) *nondecreasing* if for all $\beta \in \mathrm{Dom}(F)$ and $\alpha < \beta$ we have $F(\alpha) \leq F(\beta)$.

Remark. We dont define a decreasing ordinal function because they aren't very interesting — F can be decreasing only when $\mathrm{Dom}(F)$ is finite, since the well-ordering On does not allow infinite decreasing sequences (see Observation 1.14).

Lemma 3.2. *Increasing ordinal functions grow at least as fast as the identity. That is $F(\alpha) \geq \alpha$ for every $\alpha \in \mathrm{Dom}(F)$ for increasing F.*

Proof. For contradiction, suppose that α is the least ordinal such that $F(\alpha) < \alpha$. This means that for every $\beta < \alpha$, we have $F(\beta) \geq \beta$ (note that $\beta \in \mathrm{Dom}(F)$). Suppose $\beta = F(\alpha)$; then $F(\beta) \geq \beta = F(\alpha)$, which is a contradiction since F is increasing. □

Lemma 3.3. *If α and β are the ordinal types of the well-ordered sets A and $B \subseteq A$, then $\alpha \leq \beta$.*

Note that not necessarily $\alpha < \beta$ when $B \subsetneq A$; consider ω and $\omega \setminus \{\varnothing\}$.

Proof. Let $i_a : A \to \alpha$ and $i_b : B \to \beta$ be the isomorphisms of A and B with their types. Suppose $\beta > \alpha$ and define $f : \beta \to \alpha$ as $f = i_b^{-1} \circ i_a$. Notice that if $\gamma < \delta$, then $f(\gamma) < f(\delta)$ because both i_b^{-1} and i_a preserve order (they are order-isomorphisms), and thus f is increasing. Because $\alpha \in \mathrm{Dom}(f)$, we have that $f(\alpha) \in \mathrm{Rng}(f) = \alpha$. But this contradicts the previous lemma. □

The following is a basic property of ordinal numbers. We will not prove it, but it will be useful to keep in mind.

Lemma 3.4. *If $A \subseteq \mathrm{On}$ is a set, then $\bigcup A \in \mathrm{On}$, and in fact, $\bigcup A = \sup(A)$. We say that $\sup(A)$ is the* limit *of the sequence of ordinals A.*

Recall what you know about metric spaces, namely about closed sets and continuous functions. A subset X of a metric space is closed if each sequence in X that converges has its limit within X.

Definition 3.5. A subclass $C \subseteq \mathrm{On}$ is *closed* if, for every subset $Y \subseteq C$, we have $\sup(Y) \in C$. For an ordinal α, we say that a subset $C \subseteq \alpha$ is *closed in α* if, for every $Y \subseteq C$ satisfying $\sup(Y) < \alpha$, we have $\sup(Y) \in C$.

Observation 3.6. *If C is closed, then it has a maximum $\max(C) = \sup(C)$.*

Example. The ordinal ω is not closed because $\sup(\omega) = \omega \notin \omega$, but ω^+ is closed. In general, isolated ordinals are closed and limit are not.

You might also recall that its possible to prove that a real function is continuous if and only if the preimage of every closed set is closed. This equivalence statement is utilized in topology to define continuous functions, and we could use it here as well. But we won't need it, and we will restrict ourselves to increasing ordinal functions only, thereby making the definition simpler.

Definition 3.7. An increasing ordinal function F is *continuous* if, for every limit ordinal $\lambda \in \mathrm{Dom}(F)$, it holds that

$$F(\lambda) = \sup\{F(\alpha) \mid \alpha < \lambda\}.$$

We say that a function is *normal* if it is increasing and continuous.

Exercise 5. Show that the generalized definiton would make sense. Prove that an increasing function F is normal $\iff$ the preimage of every closed set is closed in $\mathrm{Dom}(F)$.

Example. The simplest normal function is identity. But consider the (very innocent looking) function $F(\alpha) = \alpha^+$. It is increasing, but not continuous. It fails on limit ordinals, for example $F(\omega) = \omega^+$, but

$$\sup\{F(\alpha) \mid \alpha < \omega\} = \sup\{\alpha^+ \mid \alpha < \omega\} = \sup(\omega \setminus \{\varnothing\}) = \omega.$$

Observation 3.8. *If F is a normal ordinal function and λ is a limit ordinal, then $F(\lambda)$ is a limit ordinal as well.*

Proof. Suppose that $F(\lambda) = \sup\{F(\alpha) \mid \alpha < \lambda\}$ were isolated. Then $F(\lambda) = \gamma^+$ for some $\gamma < \lambda$, which has been taken into account in the supremum. Therefore $F(\lambda) = F(\gamma)$, since F is increasing and γ is the largest ordinal smaller than λ. But this is a contradiction because $F(\gamma) < F(\lambda)$. □

Observation 3.9. *The composition $F \circ G$ of normal functions F and G is normal. This can be seen easily from the generalized definition of continous functions.*

Lemma 3.10. *If F is a normal function, then for every ordinal β, such that $F(0) \leq \beta < \sup \mathrm{Rng}(F)$, the maximum $\max\{\alpha \mid F(\alpha) \leq \beta\}$ exists.*

Intuition. For a natural number β and $f(n) = n^2$, we might consider the largest natural number α such that $F(\alpha) \leq \beta$. This α exists, it is in fact equal to $\lfloor \sqrt{\beta} \rfloor$.

Proof. Consider the closed set $[0, \beta] \coloneqq \{\alpha \mid \alpha \leq \beta\}$. It is indeed closed because $[0, \beta] = \beta^+$, which is an isolated ordinal. We will use the generalized definition of continuity (Exercise 5) and note that the preimage C of the closed set $[0, \beta]$ is closed in $\mathrm{Dom}(F)$. We would like to say that C is closed (in general). But consider $F : \omega \to \mathrm{On}$; then $\mathrm{Dom}(F)$ is not closed in On.

The bound on β will save us. Notice that there is some $\gamma \in \mathrm{Rng}(F)$ such that $\beta < \gamma \leq \sup \mathrm{Rng}(F)$. Because F is increasing, the elements of C are bounded by $F^{-1}(\gamma)$, and therefore $\sup(C) \in \mathrm{Dom}(F)$. Since C is closed in $\mathrm{Dom}(F)$, it follows that $\sup(C) \in C$ and $\sup(C) = \max(C)$, which we will denote as α. Because F is increasing, α is the largest ordinal satisfying $F(\alpha) \leq \beta$. □

Definition 3.11. An ordinal ξ is a *fixed point* of an ordinal function F if $F(\xi) = \xi$. The *class of all fixed points* of F is denoted by $K(F)$.

Example. The fixed points of the identity function are all ordinals, while the function we saw earlier, $f(\alpha) = \alpha^+$, has no fixed points. We will show that the reason is that it is not continuous.

Theorem 3.12 (About fixed points)**.** *Let $F : \mathrm{On} \to \mathrm{On}$ be a normal function.*

(i) *For every $\alpha \in \mathrm{On}$, there exists $\beta \geq \alpha$, which is a fixed point of F.*

(ii) *Concretely, consider the sequence $(\alpha_n \,|\, n \in \omega)$ defined as $\alpha_0 := \alpha$ and $\alpha_{n+1} = F(\alpha_n)$. The supremum of this sequence is the smallest of all fixed points $\xi \geq \alpha$ of F.*

(iii) *The class $K(F)$ is closed and is a proper class.*

Proof. We begin by proving (i) and (ii). Notice that $\alpha_{n+1} \geq \alpha_n$ since F grows at least as fast as the identity function. First, we show that the supremum $\beta = \sup\{\alpha_n \,|\, n \in \omega\}$ is a fixed point:

- Consider the case when $\alpha_0 < \alpha_1 < \cdots < \alpha_i = \alpha_{i+i}$ for some i; then also $\alpha_{i+2} = F(\alpha_{i+1}) = F(\alpha_i) = \alpha_i$, and thus $\beta = \alpha_i$ is a fixed point.
- Suppose the sequence never stabilizes. Since F is continuous, we have

$$\begin{aligned} F(\beta) &= \sup\{F(\gamma) \,|\, \gamma < \beta\} = \sup\{F(\alpha_n) \,|\, n \in \omega\} \\ &= \sup\{\alpha_{n+1} \,|\, n \in \omega\} = \sup\{\alpha_n \,|\, n \in \omega\} = \beta. \end{aligned}$$

Second, we show that β is the smallest fixed point larger than α. If there were a fixed point $\xi \geq \alpha$ such that $\xi < \beta$, then there would exist an index n at which $\alpha_n \leq \xi < \alpha_{n+1}$. This is because the sequence is strictly increasing, and β is its supremum. We have $\xi < \alpha_{n+1} = F(\alpha_n) \leq F(\xi)$, so ξ isn't a fixed point.

Finally, we prove (iii). We claim that $K := K(F)$ is closed. Let $C \subseteq K$ be a set; we need to show that the supremum $\beta = \sup(C)$ is a fixed point. Since F is continuous, we have

$$F(\beta) = \sup\{F(\gamma) \,|\, \gamma < \beta\} = \sup\{F(\xi) \,|\, \xi \in C\} = \sup\{\xi \,|\, \xi \in C\} = \beta.$$

To complete the proof, we show that K is a proper class. If it were a set, then by Lemma 3.4, $\sup(K)$ is an ordinal γ. We let the ordinal γ^+ take the role of α in (i) and find a new fixed point of F, larger than all those in K. $\square$

Corollary 3.13. *If $F : \mathrm{On} \to \mathrm{On}$ is a normal function, then there exists a unique order-isomorphism $J : \mathrm{On} \to K(F)$. Moreover, J is a normal function.*

Proof. The class $K(F) \subseteq \mathrm{On}$ is well-ordered, and it inherits from On the property that every proper lower part $(\leftarrow, a) \subset K(F)$ is a set (see Lemma 2.9). Exercise 3 claims that there is a unique isomorphism $J : \mathrm{On} \to K(F)$. The ordinal function J is clearly increasing (since it preserves order). It is continuous because $K(F)$ is closed. Let $\lambda \in \mathrm{On}$, we claim that $J(\lambda) = \sup\{J(\alpha) \,|\, \alpha < \lambda\}$. The set $\{J(\alpha) \,|\, \alpha < \lambda\}$ is a subset of $K(F)$, and thus its supremum lies in $K(F)$. Because λ is the smallest ordinal larger than every $\alpha < \lambda$, and J is an order-isomorphism, J has to map λ to the smallest $\kappa \in K(F)$ larger than all $\{J(\alpha) \,|\, \alpha < \lambda\}$. This is the abovementioned supremum. $\square$

3.2 Ordinal Arithmetic

At last, we will define operations such as ordinal addition and multiplication. Before proceeding further, I highly recommend watching the video [25] by Vsauce, which illustrates the concepts of constructing larger ordinals from earlier ones in a very illustrative and intuitive way.

3.2.1 Definitions and Intuition

Definition 3.14. Let α and β be ordinals. We define ordinal numbers

(a) $\alpha + \beta$ as the order type of the set $(\{0\} \times \alpha) \cup (\{1\} \times \beta)$ when ordered lexicographically,

(b) $\alpha \cdot \beta$ as the order type of the set $\beta \times \alpha$ when ordered lexicographically.

Using the popular "matchstick" representation of ordinals used in [25], $\alpha + \beta$ can be imagined as a pile of decreasing matchsticks labeled by α, followed by another pile of matchsticks labeled by β. Notice that our previous notation is consistent, as $\alpha + 1 = \alpha \cup \{\alpha\} = \alpha^+$. We first use the elements of α to label the first pile, and we need one additional ordinal to label the second pile (which contains only a single matchstick).

Notice that we are using $\beta \times \alpha$ in the definition of $\alpha \cdot \beta$. The ordinal $\alpha \cdot \beta$ can be imagined as taking multiple piles of matchsticks labeled by α and arranging them next to each other. How should the piles be arranged? In a way that we need β to label them.

With this intuition, it should not be surprising that ordinal addition and multiplication are generally not commutative. It is easy to see that $1 + \omega = \omega$ (label the first pile by 0 and the other pile by $\omega \setminus \{0\}$), but $\omega + 1 \neq \omega$. For multiplication, consider $2 \cdot \omega$, the order type of countably infinitely many copies of $\{0, 1\}$ stacked behind each other. This can be clearly labeled by ω, so $2 \cdot \omega = \omega$. But $\omega \cdot 2$ is the order type of two consecutive copies of ω. When we try to label them using ω, we use all $n \in \omega$ to label the first copy and need more ordinals for the second copy. Therefore $\omega \cdot 2 > \omega$.

This may seem troubling, as commutativity is a fundamental property of arithmetic on the natural numbers. Nevertheless, we will soon see that this familiar behavior is retained.

Observation 3.15. *For any ordinals α, β, γ and natural $n \in \omega$ it holds that*

(a) $\alpha + 0 = \alpha = 0 + \alpha, \qquad \alpha \cdot 0 = 0 = 0 \cdot \alpha, \qquad \alpha \cdot 1 = \alpha = 1 \cdot \alpha,$

(b) $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma, \qquad \alpha \cdot (\beta \cdot \gamma) = (\alpha \cdot \beta) \cdot \gamma,$

(c) $\alpha \cdot 2 = \alpha + \alpha, \qquad \alpha \cdot 3 = \alpha + \alpha + \alpha, \qquad \alpha \cdot (n + 1) = \alpha \cdot n + \alpha.$

Definition 3.16. For ordinal numbers α and β, we define α^β recursively as

(i) $\alpha^0 := 1,$

(ii) if $\beta = \gamma + 1$ is isolated, then $\alpha^\beta := \alpha^\gamma \cdot \alpha,$

(iii) if β is a limit ordinal, then $\alpha^\beta := \sup\{\alpha^\gamma \mid 0 < \gamma < \beta\}.$

To get an intuition for ordinal powers, consider the ordinal $\omega^2 = \omega \cdot \omega$. It represents multiple copies of ω arranged in the same manner as ω.

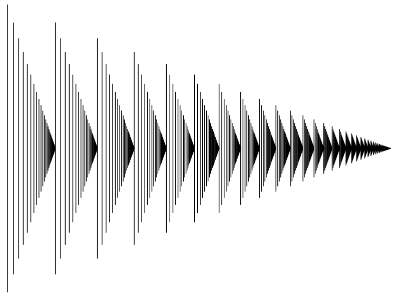

Figure 2: A representation of the ordinal ω^2. Each stick corresponds to an ordinal of the form $\omega \cdot m + n$ where m and n are natural numbers; [13].

To construct $\omega^3 = (\omega \cdot \omega) \cdot \omega$, we take multiple copies of ω^2 and arrange them in a way that requires ω to label them. If we repeat this process countably infinitely many times, we arrive at ω^ω.

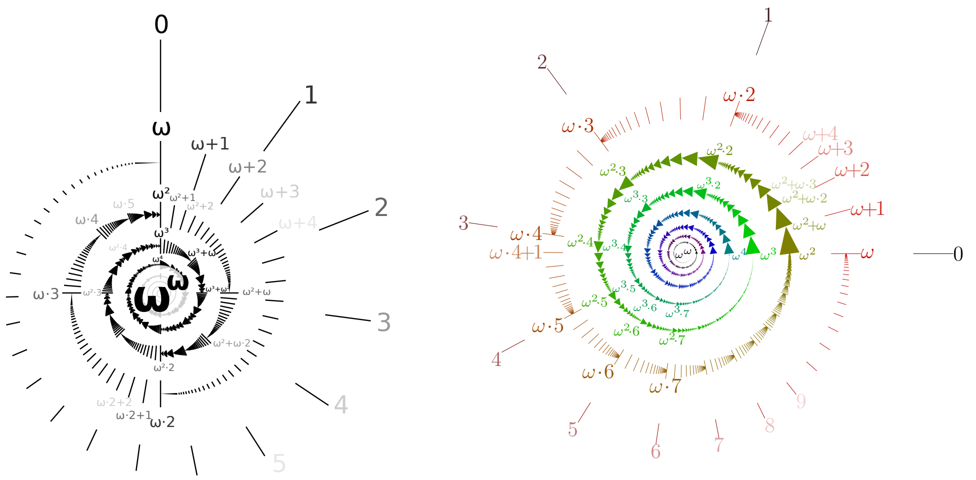

Figure 3: A spiral representation of ordinals up to ω^ω. One full turn corresponds to the mapping $f(\alpha) = \omega \cdot (1 + \alpha)$. Since ω^ω is the smallest fixed point of f, larger ordinals cannot be represented in this way; [6] and [4].

We can continue and arrive at larger and larger ordinals, such as $\omega^{(\omega^\omega)}$ or $\omega^{(\omega^{(\omega^\omega)})}$. That is a lot of parentheses, so from now on, we will write ω^{ω^ω} instead of $\omega^{(\omega^\omega)}$ and use parentheses only when we mean to say $(\omega^\omega)^\omega$.

Observation 3.17. *For any ordinals α and $\beta > 0$, it holds that*

(a) $0^0 = 1, \quad 0^\beta = 1,$

(b) $1^0 = 1, \quad 1^\beta = 1,$

(c) $\alpha^0 = 1, \quad \alpha^1 = \alpha, \quad \alpha^2 = \alpha \cdot \alpha, \quad \alpha^3 = (\alpha \cdot \alpha) \cdot \alpha.$

3.2.2 Basic Properties of Ordinal Operations

You should now have an intuition for how ordinal numbers constructed using these standard operations look. We continue by proving some of their basic properties.

Lemma 3.18 (Monotonicity of sum). *For any ordinals α and β, it holds that*

(a) $\alpha < \beta \implies \gamma + \alpha < \gamma + \beta$,

(b) $\alpha < \beta \implies \alpha + \gamma \leq \beta + \gamma$.

Proof. (a) From the definition of addition and order types, it is easy to see that $\gamma + \alpha$ is an initial segment of $\gamma + \beta$. (b) The set of ordered pairs that defines $\alpha + \gamma$ is a subset of the set of ordered pairs that defines $\beta + \gamma$. And Lemma 3.3 states that the order type of the first is at most that of the second. □

Lemma 3.19 (Monotonicity of product). *For any ordinals α, β and $\gamma > 0$, it holds that*

(a) $\alpha < \beta \implies \gamma \cdot \alpha < \gamma \cdot \beta$,

(b) $\alpha < \beta \implies \alpha \cdot \gamma \leq \beta \cdot \gamma$. *...for $\gamma = 0$ also holds*

Proof. (a) If $\alpha < \beta$, then $\alpha \times \gamma$ is an initial segment of $\beta \times \gamma$ when ordered lexicographically. (b) If $\alpha < \beta$, then $\gamma \times \alpha \subseteq \gamma \times \beta$, and the claim follows from Lemma 3.3. □

Remark. Note that the second statement in the two preceding lemmas does not, in general, hold under strict inequality. For example, $1 < 2$, but

$$1 + \omega = 2 + \omega = \omega, \qquad \text{and} \qquad 1 \cdot \omega = 2 \cdot \omega = \omega.$$

In fact, for any natural $n \in \omega$, we have that $n + \omega = \omega$ and $n \cdot \omega = \omega$.

Lemma 3.20 (Distributivity). *For any ordinals α and β_1, β_2, we have*

$$\alpha \cdot (\beta_1 + \beta_2) = \alpha \cdot \beta_1 + \alpha \cdot \beta_2.$$

That is, ordinal addition and multiplication are left-distributive. However, in general, they are not right-distributive. Meaning that for some α and β_1, β_2

$$(\beta_1 + \beta_2) \cdot \alpha \neq \beta_1 \cdot \alpha + \beta_2 \cdot \alpha$$

Proof. Left distributivity essentially states that if we arrange $\beta + \gamma$ copies of α next to each other, then it is the same as first arranging β copies of α, followed by γ copies of α. This is obviously true from how we defined addition and multiplication. However, in general, these operations are not right-distributive. Consider $(1+1) \cdot \omega \neq \omega + \omega$. On the left, we have $2 \cdot \omega$, and on the right, $\omega \cdot 2$. □

Theorem 3.21. *If m, n and k are natural numbers, then $m + n$, $m \cdot n$, and m^n are also natural numbers. Furthermore*

$$m + n = n + m, \qquad m \cdot n = n \cdot m, \qquad (m + n) \cdot k = m \cdot k + n \cdot k.$$

That is: addition and multiplication of natural numbers is commutative and right-distributive.

Proof. It is easy to see that $m+n$ and $m \cdot n$ are finite ordinals, and one can use induction on n to show that m^n is also finite.

To show that $m \cdot n = n \cdot m$, construct a bijection $f : n \times m \to m \times n$. We can use f and the lexicographic order on $n \times m$ to define a linear order $<_f$ on $m \times n$. It is well known (and it is proved in the basic set theory course) that any two linear orders on a finite set are isomorphic. Meaning that the two lexicographically ordered sets $n \times m$ and $m \times n$ are order-isomorphic; therefore, they have the same order types. Hence $m \cdot n = n \cdot m$. One can similarly show that addition commutes as well.

Right-distributivity is implied by commutativity and left-distributivity. $\square$

Lemma 3.22 (Existence of the "right" difference)**.** *If $\alpha \leq \beta$, then there is a unique ordinal ϱ such that $\alpha + \varrho = \beta$. We denote ϱ by $\beta - \alpha$.*

Intuition. Any ordinal can be extenced by a specific amount to reach any larger ordinal.

Proof. If $\alpha \leq \beta$, then $\alpha = (\leftarrow, \alpha)$ is an initial segment of β, and its complement, $\beta \setminus \alpha$, is what we might denote as $[\alpha, \rightarrow)$. If ϱ is the order type of $\beta \setminus \alpha$, then clearly $\alpha + \varrho = \beta$. The uniqueness of ϱ follows from Lemma 3.18 (a). Suppose there were $\varrho_1 < \varrho_2$ satisfying $\alpha + \varrho_1 = \beta = \alpha + \varrho_2$. But since $\varrho_1 < \varrho_2$, we have that $\alpha + \varrho_1 < \alpha + \varrho_2$. $\square$

Lemma 3.23 (Division with remainder)**.** *If $\beta > 0$, then for every ordinal α there are unique ordinals $\delta \leq \alpha$ and $\varrho < \beta$ such that $\alpha = \beta \cdot \delta + \varrho$.*

Intuition. Any ordinal α can be created by arranging multiple coppies of β in a specific way, and following this with a short tail ϱ.

Proof. Since $1 \leq \beta$, we have $\alpha \leq \beta \cdot \alpha$. If $\alpha = \beta \cdot \alpha$ (for example $\omega = 3 \cdot \omega$), choose $\delta \coloneqq \alpha$ and $\varrho \coloneqq 0$. It is not hard to show that the monotonicity of sum and product, together with left-distributivity, implies uniqueness; we will skip it.

If $\alpha < \beta \cdot \alpha$, let j be the isomorphism of the lexicographically ordered set $\alpha \times \beta$ and the ordinal $\beta \cdot \alpha$. Let $(\delta, \varrho) \in \alpha \times \beta$ be the (unique) pair mapped by j onto α. Necessarily $\delta < \alpha$ and $\varrho < \beta$. Since $\alpha \times \beta$ is ordered lexicographically, it is easy to see that $\alpha = \beta \cdot \delta + \varrho$. $\square$

Lemma 3.24 (Monotonicity of power)**.** *For any ordinals α, β, γ and $\rho > 1$, it holds that*

(a) $\alpha < \beta \implies \alpha^\gamma \leq \beta^\gamma$,

(b) $\alpha < \beta \implies \rho^\alpha < \rho^\beta$.

Proof. (a) Using transfinite induction on γ. If $\gamma = 0$, then $\alpha^\gamma = \beta^\gamma = 1$. If $\gamma = \delta + 1$ and $\alpha^\delta \leq \beta^\delta$, from the monotonicity of product we have that

$$\alpha^\gamma = \alpha^\delta \cdot \alpha \leq \beta^\delta \cdot \beta = \beta^\gamma.$$

If γ is a limit ordinal and for every $\delta < \gamma$ already $\alpha^\delta \leq \beta^\delta$, then also

$$\alpha^\gamma = \sup\{\alpha^\delta \mid 0 < \delta < \gamma\} \leq \sup\{\beta^\delta \mid 0 < \delta < \gamma\} = \beta^\gamma.$$

(b) Suppose that $\rho > 1$. It is easy to show using transfinite induction on δ that for every $\delta > 1$ it holds that $\rho^\alpha < \rho^{\alpha+\delta}$. If $\alpha < \beta$, then according to Lemma 3.22 there is a unique $\delta > 0$ satisfying $\beta = \alpha + \delta$. $\square$

Remark. Note that the first statement in the previous lemma does not, in general, hold under the strict inequality, even if $\gamma > 0$. For example, $2 < 3$, but $2^\omega = 3^\omega = \omega$. In general, if $n \in \omega$, then $n^\omega = \omega$.

Lemma 3.25 (Continuity in the second argument)**.** *The ordinal function*

(a) $F(\xi) = \alpha + \xi$ *is normal for every* $\alpha \geq 0$*,*

(b) $F(\xi) = \alpha \cdot \xi$ *is normal for every* $\alpha > 0$*,*

(c) $F(\xi) = \alpha^\xi$ *is normal for every* $\alpha > 1$*.*

Proof. All of the functions mentioned above are increasing for the specified α, since the respective operations are monotonic. We claim that they are also continuous. Let λ be a limit ordinal; we claim that $F(\lambda) = \sup\{F(\xi) \,|\, \xi < \lambda\}$. Suppose there were σ such that for all $\xi < \lambda$ we have $\sigma > F(\xi)$, but $\sigma < F(\lambda)$.

(a) $\alpha + \lambda \stackrel{?}{=} \sup\{\alpha + \xi \,|\, \xi < \lambda\}$: Lemma 3.22 claims that there is a unique ϱ such that $\sigma = \alpha + \varrho$. Because $\sigma < F(\lambda)$ we have that $\alpha + \varrho < \alpha + \lambda$ and thus $\varrho < \lambda$. It should hold that $\sigma > F(\varrho)$, but $\sigma = \alpha + \varrho = F(\varrho)$.

(b) $\alpha \cdot \lambda \stackrel{?}{=} \sup\{\alpha \cdot \xi \,|\, \xi < \lambda\}$: Lemma 3.23 claims the existence of unique ordinals δ and $\varrho < \alpha$ such that $\sigma = \alpha \cdot \delta + \varrho$. Since $\sigma < \alpha \cdot \lambda$ we have that $\delta < \lambda$ (monotonicity) and also $\delta + 1 < \lambda$ (λ is limit). Combining these we get

$$\sigma = \alpha \cdot \delta + \varrho < \alpha \cdot \delta + \alpha = \alpha \cdot (\delta + 1) = F(\delta + 1).$$

But we assumed that $\sigma > F(\xi)$ for all $\xi < \lambda$.

(c) $\alpha^\lambda \stackrel{?}{=} \sup\{\alpha^\xi \,|\, \xi < \lambda\}$: This just the definition of α^λ for $\alpha > 1$. □

Lemma 3.26 (Addition and multiplication in the exponent)**.** *For any ordinals* α, β *and* γ*, it holds that*

(a) $\alpha^{\beta+\gamma} = \alpha^\beta \cdot \alpha^\gamma$,

(b) $(\alpha^\beta)^\gamma = \alpha^{\beta \cdot \gamma}$

Proof. (a) Trivially holds for $\alpha \leq 1$. Suppose $\alpha > 1$, we will use transfinite induction on γ. If $\gamma = 0$, there is nothing to prove. If $\gamma = \delta + 1$ is isolated, then $\beta + \gamma = (\beta + \delta) + 1$, and the statement follows from the induction hypothesis:

$$\alpha^{(\beta+\delta)+1} = \alpha^{(\beta+\delta)} \cdot \alpha^1 = \alpha^\beta \cdot \alpha^\delta \cdot \alpha = \alpha^\beta \cdot \alpha^{\delta+1} = \alpha^\beta \cdot \alpha^\gamma.$$

Finally, if γ is a limit ordinal and the statement holds for all $\gamma < \delta$, then

$$\begin{aligned}
\alpha^\beta \cdot \alpha^\gamma &= \sup\{\alpha^\beta \cdot \xi \,|\, \xi < \alpha^\gamma\} && \ldots F(\xi) = \alpha^\beta \cdot \xi \text{ is normal} \\
&= \sup\{\alpha^\beta \cdot \alpha^\delta \,|\, 0 < \delta < \gamma\} && \ldots \alpha^\gamma = \sup\{\alpha^\delta \,|\, 0 < \delta < \gamma\} \\
&= \sup\{\alpha^{\beta+\delta} \,|\, 0 < \delta < \gamma\} && \ldots \text{induction hypothesis} \\
&= \sup\{\alpha^\varepsilon \,|\, 0 < \varepsilon < \beta + \gamma\} && \ldots F(\xi) = \beta + \xi \text{ is normal} \\
&= \alpha^{\beta+\gamma}.
\end{aligned}$$

(b) Suppose that $\beta, \gamma \neq 0$ and $\alpha > 1$, otherwise, it trivially holds. We will again use transfinite induction on γ. If $\gamma = 0$, then it holds. If $\gamma = \delta + 1$ is isolated, then the statement follows from (a) and the induction hypothesis:

$$(\alpha^\beta)^{\delta+1} = (\alpha^\beta)^\delta \cdot (\alpha^\beta)^1 = \alpha^{\beta\cdot\delta} \cdot \alpha^\beta = \alpha^{\beta\cdot\delta+\beta} = \alpha^{\beta\cdot(\delta+1)} = \alpha^{\beta\cdot\gamma}.$$

Finally, if γ is a limit ordinal, then $\beta \cdot \gamma$ is also limit and

$$\begin{aligned}
(\alpha^\beta)^\gamma &= \sup\{(\alpha^\beta)^\delta \mid 0 < \delta < \gamma\} && \ldots \gamma \text{ is limit} \\
&= \sup\{\alpha^{\beta\cdot\delta} \mid 0 < \delta < \gamma\} && \ldots \text{induction hypothesis} \\
&= \sup\{\alpha^\varepsilon \mid 0 < \varepsilon < \beta\cdot\gamma\} && \ldots F(\xi) = \beta \cdot \xi \text{ is normal} \\
&= \alpha^{\beta\cdot\gamma}.
\end{aligned}$$

The last equality holds because $\beta \cdot \gamma$ is a limit ordinal. □

3.2.3 Ordinal Equations and Power Expansions

Example. Suppose we want to find all ξ and β satisfying $\xi + \beta = \omega$. Lemma 3.22 claims that $\xi \leq \omega$ and $\beta = \omega - \xi$. Suppose $\xi = \omega$, then $\beta = 0$. If $\xi = n$ is a natural number, then $\beta = \omega - n = \omega$. We conclude that β can attain only two different values.

Proposition 3.27. *Let α be an ordinal and consider the equation $\xi + \beta = \alpha$. The set of solutions (ξ, β) contains only finitely many distinct values of β.*

Proof. Suppose that for some α, there are infinitely many distinct values of β in the solution set. Let $(\xi_n, \beta_n)_{n\in\omega}$ be a sequence of solutions such that $\beta_n < \beta_{n+1}$ for all n. Since $\xi_n + \beta_n = \xi_{n+1} + \beta_{n+1}$, from the monotonicity of sum we have that $\xi_n > \xi_{n+1}$ for all $n \in \omega$. We have constructed an infinite strictly decreasing sequence, which is impossible since On is well-ordered. □

We are able to express any natural number n as an expansion of powers of any base $b > 1$. We will prove that a similar statement holds for ordinal numbers too. A base of special importance is ω (as it is the first transfinite ordinal), and the expansion of α over ω is called its *Cantor normal form*; however, an expansion is possible over any base $\beta > 1$.

Lemma 3.28. *If $k, m_0, m_1, \ldots, m_k$ are natural numbers and $\gamma_0, \gamma_1, \ldots, \gamma_k > \delta$ are ordinals, then*

$$\omega^\delta > \omega^{\gamma_0} \cdot m_0 + \omega^{\gamma_1} \cdot m_1 + \cdots + \omega^{\gamma_k} \cdot m_k.$$

Proof. Let m be the largest among all m_i, and γ be the largest among all γ_i. Then $\omega^\gamma \cdot m \cdot k$ is an upper bound of the sum on the right side of the equation. We assumed that $\delta \geq \gamma + 1$, so $\omega^\delta \geq \omega^{\gamma+1} > \omega^\gamma \cdot m \cdot k$. □

Theorem 3.29 (Expansion over ω)**.** *For any $\alpha > 0$ there are unique natural numbers $k, m_0, m_1, \ldots, m_k \neq 0$ and ordinals $\gamma_0 > \gamma_1 > \cdots > \gamma_k$ which satisfy*

$$\alpha = \omega^{\gamma_0} \cdot m_0 + \omega^{\gamma_1} \cdot m_1 + \cdots + \omega^{\gamma_k} \cdot m_k. \tag{3.1}$$

The sum on the right side of the equation is called the Cantor normal form of α.

Furthermore, if

$$\beta = \omega^{\delta_0} \cdot n_0 + \omega^{\delta_1} \cdot n_1 + \cdots + \omega^{\delta_l} \cdot n_l \tag{3.2}$$

is the Cantor normal form of an ordinal β, then $\beta > \alpha$ if and only if one of the two following cases occurs:

(a) *$l > k$ and the first k terms of β are idential to those of α. For example: $\alpha = \omega^2 + \omega \cdot 2$ and $\beta = \omega^2 + \omega \cdot 2 + 3$.*

(b) *there exists an index $i \leq \min(k, l)$ at which (γ_i, m_i) and (δ_i, n_i) differ, and for the smallest such index i either $\delta_i > \gamma_i$, or $\delta_i = \gamma_i$ and $n_i > m_i$. For example $\alpha = \omega^2 + \omega \cdot 2$ and $\beta = \omega^2 \cdot 5 + \omega \cdot 2$.*

Proof. We prove the first part by transfinite induction on α. The CNF of $\alpha = 1$ is $\alpha = \omega^0 \cdot 1$. Suppose $\alpha > 0$ and that every nonzero $\beta < \alpha$ has a unique CNF. The ordinal function $\gamma \mapsto \omega^\gamma$ is normal, so according to Lemma 3.10, there exists a maximal ordinal γ such that $\omega^\gamma \leq \alpha$. Similarly, from the normality of product in the second argument follows the existence of a maximal ordinal δ such that $\omega^\gamma \cdot \delta \leq \alpha$. Also, $\delta < \omega$, since $\omega^{\gamma+1} = \omega^\gamma \cdot \omega > \alpha$, which contradicts the choice of γ. If $\omega^\gamma \cdot \delta = \alpha$, then the uniqueness of this expansion follows from Lemma 3.28. If $\omega^\gamma \cdot \delta < \alpha$, then there exists a unique ordinal $\beta = \alpha - \omega^\gamma \cdot \delta$ such that $\omega^\gamma \cdot \delta + \beta = \alpha$. Note that $\beta < \omega^\gamma$; otherwise, we get $\omega^\gamma \cdot \delta + \beta \geq \omega^\gamma \cdot (\delta + 1)$, which contradicts the choice of δ.

To find the CNF on α, let

$$\beta = \omega^{\gamma_1} \cdot m_1 + \omega^{\gamma_2} \cdot m_2 + \cdots + \omega^{\gamma_k} \cdot m_k$$

be the CNF of β. Define $\gamma_0 \coloneqq \gamma$ and $m_0 \coloneqq \delta$. Then $\gamma_0 > \gamma_1$, and (3.1) is the CNF of α. The uniqueness of this expansion follows from Lemma 3.28 and the unique choice of β.

Next, we prove the second part of the theorem. Suppose that the ordinals α and β have Cantor normal forms (3.1) and (3.2). If (a) holds, $\beta > \alpha$ because the trailing terms in the expansion of β are nonzero. Suppose that (b) holds and that i is the least index at which the two expansions differ. If $\delta_i > \gamma_i$, then $\beta > \alpha$ from Lemma 3.28. If $\delta_i = \gamma_i$ and $n_i > m_i$, then $n_i \geq m_i + 1$ and

$$\omega^{\delta_i} \cdot n_i \geq \omega^{\gamma_i} \cdot m_i + \omega^{\gamma_i}.$$

Lemma 3.28 claims that the second summand on the right (ω^{γ_i}) is a strict upper bound of the remaining summands in (3.1), the expansion of α; thus $\beta > \alpha$.

All that remains is to prove the reverse implication. If $\beta > \alpha$, then their Cantor normal forms (3.1) and (3.2) have to differ. Either the CNF of one of the ordinals is the same as the beginning of the CNF of the other, or there exists an index at which they differ. We can use the already proven implication to show that the only two possible cases are (a) and (b). □

Corollary 3.30 (Alternative expansions)**.** *For any $\alpha > 0$, it holds that*

(a) *there is a unique natural number $l > 0$ and unique ordinals $\gamma_0 \geq \gamma_1 \geq \cdots \geq \gamma_l$ which satisfy*

$$\alpha = \omega^{\gamma_0} + \omega^{\gamma_1} + \cdots + \omega^{\gamma_l},$$

(b) *there are unique ordinals β and γ such that*

$$\alpha = \omega^\gamma \cdot (\beta + 1).$$

Proof. (a) For any ordinal γ and natural m, is the ordinal number $\omega^\gamma \cdot m$ equal to the sum of m summands of the form ω^γ. We obtain the expansion in (a) by expressing each term in the CNF of α in this expanded form.

(b) If α has CNF (3.1), we let $\gamma = \gamma_k$. Then for all $i \leq k$ is $\gamma_i = \gamma + \delta_i$ for $\delta_i = \gamma_i - \gamma$. From the properties of exponents and left-distributivity, we get

$$\alpha = \omega^\gamma \cdot \big(\omega^{\delta_0} \cdot m_0 + \omega^{\delta_1} \cdot m_1 + \cdots + \omega^0 \cdot m_k\big).$$

The parentheses on the right contain an isolated ordinal $\beta + 1$, because m_k is a nonzero natural number. The uniqueness of the ordinals γ and β follows from the uniqueness of the CNF of α. □

Theorem 3.31 (Expansion over any base)**.** *The choice of ω as a base in Theorem 3.29 was arbitrary; the same holds for any ordinal base $\beta > 1$. We just need to restrict the nonzero coefficients $m_0, m_1, \dots, m_k$ to be smaller than β.*

Proof. We did not use any special properties of ω in the proof of Theorem 3.29, so we only need to modify Lemma 3.28. If we slightly change its claim, only for decreasing exponents $\gamma_0 > \gamma_1 > \cdots > \gamma_k$ and coefficients $m_i < \beta$, we can prove it by transfinite induction on γ_i. □

Remark. If we restrict ourselves only to natural numbers, we obtain the familiar theorem about expanding natural numbers using powers of a base $b > 1$.

3.2.4 Escaping ω and the Epsilon Numbers

We saw earlier that for all natural numbers n, it holds that $n+\omega = n\cdot\omega = n^\omega = \omega$. We also proved that the functions corresponding to these basic operations,

$$A_n(\xi) = n + \xi, \qquad M_n(\xi) = n \cdot \xi, \qquad E_n(\xi) = n^\xi,$$

are normal. Theorem 3.12 claims that each of them has infinitely many fixed points. It is easy to see that no (nonzero) natural number is a fixed point, and above we have observed that ω is a fixed point of all of them. It is, in fact, the smallest (nonzero) fixed point. Notice that this makes intuitive sense. When restricted to natural numbers, these are all fast growing functions ($A \ll M \ll E$), so we need a new concept (countable infinity) to find a fixed point.

Now consider what would happen if we replaced n with ω and tried to find a (nonzero) fixed point of these new ω-functions. Theorem 3.12 claims that the smallest such fixed points are:

- $F_A = \sup\{0, \omega, \omega+\omega, \omega+\omega+\omega, \omega\cdot 4, \omega\cdot 5, \dots\} = \omega\cdot\omega$,
- $F_M = \sup\big\{1, \omega, \omega\cdot\omega, \omega\cdot\omega\cdot\omega, \omega^4, \omega^5, \dots\big\} = \omega^\omega$,
- $F_E = \sup\big\{1, \omega, \omega^\omega, \omega^{\omega^\omega}, \omega^{\omega^{\omega^\omega}}, \dots\big\}$, and we no longer have notation to describe this number; we will denote it as ε_0.

The question is: did we escape the countable infinity represented by ω? No, we will show that all of these numbers are, in fact, still countable. Nonetheless, we have stumbled upon something important. The last number, $\varepsilon_0 = \omega^{\varepsilon_0}$, is closely connected to Peano arithmetic, and we will soon encounter it when proving Goodsteins theorem. It also gives rise to an entire class of ordinals called the *epsilon numbers.*

Definition 3.32. An ordinal ξ is an ε-*number* if it is a fixed point of the normal function $F(\xi) = \omega^\xi$. That is, if $\xi = \omega^\xi$. Corollary 3.13 asserts the existence of an isomorphism $J : \mathrm{On} \to \{\varepsilon \,|\, \varepsilon = F(\varepsilon)\}$; we denote by ε_β the ordinal $J(\beta)$.

Proposition 3.33. *For any ordinal β, it holds that*

(i) $\varepsilon_0 = \sup\left\{1, \omega, \omega^\omega, \omega^{\omega^\omega}, \omega^{\omega^{\omega^\omega}}, \dots\right\}$,

(ii) $\varepsilon_{\beta+1} = \sup\left\{1, \varepsilon_\beta, {\varepsilon_\beta}^{\varepsilon_\beta}, {\varepsilon_\beta}^{{\varepsilon_\beta}^{\varepsilon_\beta}}, {\varepsilon_\beta}^{{\varepsilon_\beta}^{{\varepsilon_\beta}^{\varepsilon_\beta}}}, \dots\right\}$,

(iii) $\varepsilon_\beta = \sup\{\varepsilon_\delta \,|\, \delta < \beta\}$, *whenever β is a limit ordinal.*

Proof. We prove the theorem by transfinite induction on β. (i) is the definition of ε_0, and (iii) is true because the isomorphism J is a normal function (see Corollary 3.13). (ii) Following Theorem 3.12, we know that $\varepsilon_{\beta+1}$ is the limit of the sequence

$$\varepsilon_\beta + 1, \omega^{\varepsilon_\beta+1}, \omega^{\omega^{\varepsilon_\beta+1}}, \omega^{\omega^{\omega^{\varepsilon_\beta+1}}}, \dots$$

Let β_n denote the element with index $n \in \omega$. Define a different sequence for $n \geq 2$ as $\beta'_2 \coloneqq \varepsilon_\beta^{\ \omega}$ and $\beta'_{n+1} \coloneqq \varepsilon_\beta^{\ \beta'_n}$. Clearly

$$\sup\{\beta'_n \,|\, n \geq 2\} = \sup\left\{1, \varepsilon_\beta, {\varepsilon_\beta}^{\varepsilon_\beta}, {\varepsilon_\beta}^{{\varepsilon_\beta}^{\varepsilon_\beta}}, {\varepsilon_\beta}^{{\varepsilon_\beta}^{{\varepsilon_\beta}^{\varepsilon_\beta}}}, \dots\right\}.$$

We will use induction on n to show $\beta_n = \beta'_n$ for all $n \geq 2$:

$$\beta_1 = \omega^{\varepsilon_\beta+1} = \omega^{\varepsilon_\beta} \cdot \omega = \varepsilon_\beta \cdot \omega$$
$$\beta_2 = \omega^{\omega^{\varepsilon_\beta+1}} = \omega^{(\varepsilon_\beta \cdot \omega)} = (\omega^{\varepsilon_\beta})^\omega = \varepsilon_\beta^{\ \omega} = \beta'_2$$
$$\beta_3 = \omega^{\omega^{\omega^{\varepsilon_\beta+1}}} = \omega^{\varepsilon_\beta^{\ \omega}} = \omega^{\varepsilon_\beta^{\ 1+\omega}} = \omega^{\varepsilon_\beta \cdot \varepsilon_\beta^{\ \omega}} = (\omega^{\varepsilon_\beta})^{\varepsilon_\beta^{\ \omega}} = \varepsilon_\beta^{\ \varepsilon_\beta^{\ \omega}} = \beta'_3$$
$$\beta_{n+2} = \omega^{\beta_{n+1}} = \omega^{\beta'_{n+1}} = \omega^{\varepsilon_\beta^{\ \beta'_n}} = \omega^{\varepsilon_\beta^{\ 1+\beta'_n}} = \omega^{\varepsilon_\beta \cdot \varepsilon_\beta^{\ \beta'_n}} = (\omega^{\varepsilon_\beta})^{\beta'_{n+1}} = {\varepsilon_\beta}^{\beta'_{n+1}} = \beta'_{n+2} \qquad \square$$

Lemma 3.34. *A countable union of countable ordinals is countable (in* ZF*).*

Remark. If we assumed AC, then a similar proof would imply that a countable union of arbitrary countable sets is countable.

Proof. Let $\{\alpha_n \,|\, n < \omega\}$ be a countable set of countable ordinals, WLOG $\alpha_0 < \alpha_1 < \alpha_2 < \cdots$. Denote

$$S \coloneqq \bigcup_{n<\omega} \alpha_n = \sup\{\alpha_n \,|\, n < \omega\}.$$

We will define an injection $g : S \to \omega \times \omega$ (here, $\omega \times \omega$ is countable[4]). Since each α_n is countable, there exists an injection $j_n : \alpha_n \to \omega$. For any ordinal $\xi \in S$, define

$$n_\xi \coloneqq \min\{n \in \omega \mid \xi < \alpha_n\}.$$

This number indicates in which α_n does ξ first appear in. Notice that more ordinals ξ can have the same number n_ξ, but that $j_{n_\xi}(\xi)$ gives us a unique identifier of ξ among these ordinals (since j_{n_ξ} is injective). Hence, we can define an injection $g : \xi \mapsto (n_\xi, j_{n_\xi}(\xi))$. □

Lemma 3.35. *The epsilon number ε_0 is countable.*

Proof. We will show that all ordinals $\alpha < \varepsilon_0$ are countable using transfinite induction on α up to ε_0. If $\alpha = 0$, trivially holds. If $\alpha = \beta + 1$ is isolated, then (by the induction hypothesis) β is countable. If β is finite, then α is also finite. If $\beta \approx \omega$, then it is easy to create an injection $\alpha \to \beta$.

Finally, suppose that α is a limit ordinal with Cantor normal form

$$\alpha = \omega^{\gamma_0} \cdot m_0 + \omega^{\gamma_1} \cdot m_1 + \cdots + \omega^{\gamma_k} \cdot m_k.$$

Because $\alpha < \varepsilon_0$ (which is the first fixed point of ω^ξ), we know that $\gamma_0 < \alpha$. Now, we can show that α is countable using our induction hypothesis.

- The exponents γ_i are countable since $\alpha > \gamma_0 > \gamma_1 > \cdots$.
- The ordinals ω^{γ_i} are countable as well. Indeed, notice that Lemma 3.10 implies that ω^{γ_0} is the set of all ordinals that have CNF with exponents strictly smaller than γ_0. That is, any ordinal $\beta < \omega^{\gamma_0}$ has a CNF

  $$\beta = \omega^{\delta_0} \cdot d_0 + \omega^{\delta_1} \cdot d_1 + \cdots + \omega^{\delta_l} \cdot m_l,$$

  where all $\delta_j < \gamma_0$. The set of all possible exponents is thus γ_0, and the set of all possible coefficients is ω, both of which are countable. Hence, the set of all possible pairs (δ_j, d_j) is $\gamma_0 \times \omega$, which is countable as well. Any $\beta \in \omega^{\gamma_0}$ is representable by a unique finite sequence of these pairs (we have an injection $\omega^{\gamma_0} \to$ finite sequences), and because the set of all finite sequences of elements from a countable set is countable, we arrive at the conclusion that ω^{γ_0} is countable. Because all other γ_i are smaller than γ_0, we get that each ω^{γ_i} is countable as well.
- The terms $\omega^{\gamma_i} \cdot m_i$ are countable because m_i is countable and the ordinal $\omega^{\gamma_i} \cdot m_i$ has the same size as the Cartesian product $\omega^{\gamma_i} \times m_i$ of two countable sets, which is countable.
- Finally, α is countable because it is a finite sum of terms of the form $\omega^{\gamma_i} \cdot m_i$, each of which is countable. Again, we have that α is a finite union of countable sets and is, therefore, countable as well.

[4]Prove that the Cartesian product of finitely many countable sets is countable. Hint: prime numbers might help.

We have shown that every $\alpha < \varepsilon_0$ is countable. Since

$$\varepsilon_0 = \sup\{\alpha_n \mid n < \omega\} = \bigcup_{n<\omega} \alpha_n,$$

where $\alpha_0 = 1$ and $\alpha_{n+1} = \omega^{\alpha_n}$, is a countable union of countable ordinals, the last lemma asserts that ε_0 is countable. $\square$

We have shown that ε_0 is countable, but perhaps a larger epsilon number will prove to be uncountable?

Theorem 3.36. *The epsilon number ε_β is countable $\iff$ β is countable.*

Proof. We first prove '$\Leftarrow$' by transfinite induction on the countable ordinal β. The base case $\beta = 0$ has been verified by the previous lemma. Suppose the claim holds for a countable ordinal β, that is, ε_β is countable, and we want to show that $\varepsilon_{\beta+1}$ is countable as well. Theorem 3.12 claims that $\varepsilon_{\beta+1}$ is the limit of the sequence $\alpha_0 = \varepsilon_\beta + 1$ and $\alpha_{n+1} = \omega^{\alpha_n}$. Crucially, $\varepsilon_{\beta+1}$ is the first fixed point of ω^ξ larger than ε_β, meaning that for all ordinals α such that $\varepsilon_\beta < \alpha < \varepsilon_{\beta+1}$, we have that $\alpha < \omega^\alpha$.

We already know that all $\alpha \leq \varepsilon_\beta$ are countable; we can use the previous observation to show that all $\varepsilon_\beta < \alpha < \varepsilon_{\beta+1}$ are also countable by replicating the proof we did earlier for ε_0. Alternatively, we could modify the proof by replacing all CNF expansions with base ε_β expansions; this way, we can prove the countability of all $\alpha < \varepsilon_{\beta+1}$ directly. Either way, we can conclude that all α_n are countable and that $\varepsilon_{\beta+1}$, which is a countable union of these countable ordinals, is also countable.

Finally, assume that β is a countable limit ordinal. Proposition 3.33 claims that

$$\varepsilon_\beta = \sup\{\varepsilon_\delta \mid \delta < \beta\} = \bigcup\{\varepsilon_\delta \mid \delta < \beta\}.$$

Since β is countable, this is a countable union of countable ordinals (induction hypothesis), so it is countable as well.

We prove '$\Rightarrow$' by contraposition. Suppose that β is uncountable; we will show that ε_β is also uncountable. Let $J : \alpha \mapsto \varepsilon_\alpha$ be the isomorphism from the definition of epsilon numbers and consider the restriction $J \restriction \beta$. This is an injection $\beta \to \varepsilon_\beta$. Since β is uncountable, ε_β has to be uncountable too. $\square$

The theorem we have just proven places us in a difficult position. Does an uncountable ordinal even exist? If we assume the axiom of choice, then it is fairly easy to find one: just well-order the uncountable set $\mathcal{P}(\omega)$ and take its order type. Finding one in ZF seems to be much more difficult.

We know that combining countable ordinals using the standard operations defined above produces more countable ordinals. The best tool for constructing large ordinals we currently have are the epsilon numbers (and, moreover, Theorem 3.12 implies that for any ordinal, there is a larger epsilon number), but it seems like they will not help us either. Consider the sequence

$$\gamma_0 = \varepsilon_0,\ \gamma_{n+1} = \varepsilon_{\gamma_n}. \qquad \text{That is,} \qquad \varepsilon_0, \varepsilon_{\varepsilon_0}, \varepsilon_{\varepsilon_{\varepsilon_0}}, \varepsilon_{\varepsilon_{\varepsilon_{\varepsilon_0}}}, \ldots$$

The largest number we can currently construct is the limit of this sequence; the first fixed point of the epsilon function $\xi \mapsto \varepsilon_\xi$, a number denoted as ζ_0. However, this number, while enormously large, is still countable. The reason is that all the terms γ_n are countable (by induction and the previous theorem); hence, ζ_0 is a countable union of countable ordinals and is thus also countable.

We could define *zeta* (ζ) *numbers* in a similar fashion to how we defined epsilon numbers; however, for the same reasons that epsilon numbers with countable indices are countable, we would arrive at the conclusion that any zeta number with a countable index is still countable.

We could even create an entire hierarchy of these special fixed-point numbers. The bottom tier would be $\varphi_0(\beta) = \omega^\beta$; the second tier would be the epsilon numbers $\varphi_1(\beta) = \varepsilon_\beta$; the third tier would be the zeta numbers $\varphi_2(\beta) = \zeta_\beta$; the next one would be the so-called *eta numbers* $\varphi_3(\beta) = \eta_\beta$, and so on. The tiers are defined in such a way that the values of φ_{n+1} are the fixed points of φ_n. We could now consider the ordinal

$$\Gamma_0 \coloneqq \sup\{\varphi_n(0) \mid n < \omega\}.$$

However, this ordinal is *still countable*, as it is a countable union of countable ordinals. We could even generalize φ_n to all ordinals α, and we would notice that as long as α is countable, the ordinals in tier φ_α are also countable.

Theorem 3.37. *This argument demonstrates a profound concept in set theory: one cannot reach uncountability by starting from ω and applying operations such as addition, multiplication, exponentiation, finding fixed points of normal functions, or taking suprema any countable number of times.*

Remark. The hierarchy of ordinals that we have just described is called the *Veblen Hierarchy*, the function φ_α is called the *Veblen function*, and the ordinal Γ_0, known as the *Feferman–Schütte ordinal*, is one of the most famous ordinals in all of logic. We will attempt to provide an idea of why in the next section.

Does that mean that all hope is lost and there are no uncountable ordinals? Thankfully, no. The following theorem gives us a way out.

Theorem 3.38 (Hartogs, 1915)**.** *For any set x, there exists an ordinal η such that there is no injection $\eta \to x$. The least such η is called the* Harotgs number *of x.*

Proof (cf. [9]*).* Consider the set[5]

$$\mathcal{W} = \{(A, <_R) \mid A \subseteq x \text{ and } <_R \text{ is a well-ordering of } A\}.$$

We can use replacement to construct the set

$$S = \{\alpha \in \mathrm{On} \mid \text{there exists } (A, <_R) \in \mathcal{W} \text{ order-isomorphic to } \alpha\}$$

by assigning to each $(A, <_R)$ its order type.

But this set is exactly the Hartogs number of x. Notice that S is transitive: if $\alpha \in S$ and $\gamma < \alpha$, then $\alpha \in S$ as well. A transitive set of ordinals is again an ordinal, so S is an ordinal number η. Furthermore, there is no injection from η into x, because if there were, then we would get the contradiction that $\eta \in \eta$. And finally, η is the least such ordinal. If $\alpha < \eta$, then also $\alpha \in \eta$, and there is an injection $\alpha \to x$. $\square$

[5] Why is this a set?

Remark. It is crucial to note that the theorem we have just proved, which gives us the Hartogs number as a von Neumann ordinal, is more powerful than Hartogs's original 1915 result. Hartogs, working in Z (proposed in 1908 by Zerlemlo, containing the axioms of ZF except Replacement), only proved the existence of a well-ordered set that could not be injected into x; but he did not—and could not—show it was a von Neumann ordinal. The general theorem that "every well-ordered set is isomorphic to a unique von Neumann ordinal" is itself not provable in Z and requires replacement (see proof of Theorem 2.13). We will now use this modern, replacement-based construction to construct an uncountable ordinal as the Hartogs number of ω. It is this very step, guaranteeing that the collection of all countable ordinals is a set, that fails in Z and was a key motivation for Fraenkel and Skolem to propose the axiom of replacement in 1922.

Notice that the theorem does not say that $x \prec \eta$, because this does not in general hold without AC. However, if x is well-ordered, then it has an order type α, and we can compare α with η.

This allows us to access an uncountable ordinal. Let ω_1 be the Hartogs number of ω. Either $\omega_1 \leq \omega$, which is impossible since there would be an injection $\omega_1 \to \omega$, or $\omega_1 > \omega$, which gives us a way to construct an injection $\omega \to \omega_1$; therefore, $\omega \preceq \omega_1$. But since $\omega_1 \not\preceq \omega$, the Cantor–Berstein theorem implies that $\omega \prec \omega_1$. By the definition of Hartogs numbers, ω_1 is the least ordinal with this property, meaning that it is the *first uncountable ordinal*, and we can write

$$\omega_1 = \{\alpha \in \mathrm{On} \mid \alpha \preceq \omega\}$$

It is hard to describe just how absurdly large ω_1 is. The entire vast, complex, mind-boggling hierarchy of ordinals described by φ_α (for countable α) is still just a tiny, countable speck at the absolute "bottom" of the ordinal line from the perspective of ω_1.

Observation 3.39. *The ordinal ω_1 is an epsilon number. In fact, $\omega_1 = \varepsilon_{\omega_1}$.*

Proof. Since ω_1 is a limit ordinal, Proposition 3.33 claims that

$$\varepsilon_{\omega_1} = \sup\{\varepsilon_\delta \mid \delta < \omega_1\}.$$

It is not hard to see that this is equal to $\sup\{\delta \mid \delta < \omega_1\}$, which is, by definition, equal to ω_1. $\square$

Realize that there was nothing special about the choice of ω. We can apply the same process to ω_1 to get ω_2, and continue doing this to construct larger and larger ordinals (in the sense of cardinality).

Definition 3.40. For an ordinal α we define ω_α as

(i) $\omega_0 \coloneqq \omega$,

(ii) if $\alpha = \beta + 1$ is isolated, then ω_α is the Hartogs number of ω_β,

(iii) if α is a limit ordinal, then $\omega_\alpha \coloneqq \sup\{\omega_\delta \mid \delta < \alpha\}$.

Observation 3.41. *The number ω_α is the first ordinal that is larger (in the sense of cardinality) than all previous ω-numbers.*

This definition foreshadows the section about cardinal numbers, where we will encounter these omega numbers again and explore their properties in depth.

Hartogs numbers also allow us to finally prove that the trichotomy principle implies AC. In fact, this was the original motivation behind Hartogs's theorem.

Theorem 3.42. *The trichotomy principle implies the well-ordering principle.*

Proof. Let x be an arbitrary set, and let η be its Hartogs number. Apply the trichotomy principle to x and η. Exactly one of the following holds:

(a) $x \preceq \eta$, there is an injection $x \to \eta$, or

(b) $\eta \preceq x$, there is an injection $\eta \to x$.

The second case is impossible due to the defining property of η. Hence, there exists an injection $f : x \to \eta$. We can now well-order x by inheriting the order of η by f. $\square$

3.3 Peano Arithmetic

To understand this section, the reader should be familiar with the basic notions of logic, including concepts such as language, theory, model, etc. The lecture notes [3] for the course NAIL062 provide a good foundation.

3.3.1 Peano Axioms

Peano Arithmetic, denoted PA, is the standard axiomatic theory of the natural numbers. In ZFC, we have encountered the set of natural numbers, ω, constructed as the set of finite von Neumann ordinals. This is no coincidence; the set ω, together with the restrictions of operations of ordinal arithmetic to ω, serves as the *standard model* for PA, denoted by $\mathcal{N}$.

Our study of ordinal arithmetic in Section 3.2 has already established that these operations, when restricted to finite ordinals, are commutative and satisfy all the familiar properties of elementary arithmetic. The axioms of PA, therefore, can be seen as a precise, first-order logic attempt to capture the properties of this standard model.

Definition 3.43 (PA, [3])**.** The language of PA is $\mathcal{L}_{PA} = \langle 0, S, +, \cdot, \leq \rangle$ with equality. The base axioms of PA are the following formulas:

$$\begin{array}{ll}
\neg Sx = 0 & x \cdot 0 = 0 \\
Sx = Sy \implies x = y & x \cdot Sy = x \cdot y + x \\
x + 0 = x & \neg x = 0 \implies (\exists y)(x = Sy) \\
x + Sy = S(x + y) & x \leq y \iff (\exists z)(z + x = y)
\end{array}$$

These axioms alone yield the much weaker *Robinson Arithmetic* (Q). It cannot prove, for example, the commutativity or associativity of addition or multiplication, or the transitivity of order. To obtain PA, we need to add the *Axiom Schema of Induction.* That is, for each $\mathcal{L}_{PA}$-formula $\varphi(x, \vec{y})$, the following axiom is added:

$$\big(\varphi(0, \vec{y}) \wedge (\forall x)(\varphi(x, \vec{y}) \to \varphi(Sx, \vec{y}))\big) \implies (\forall x)\varphi(x, \vec{y}) \tag{3.3}$$

Remark. The last axiom schema should seem similar to the induction principle on ω from set theory:

$$(\forall X \subseteq \omega)\Big(\big(0 \in X \wedge (\forall x)(x \in X \Rightarrow x \cup \{x\} \in X)\big) \implies X = \omega\Big).$$

However, the axiom schema of induction is a weaker version, as it is a first-order logic attempt to simulate a second-order logic axiom with an axiom schema. The familiar induction principle could be expressed with the following second-order $\mathcal{L}_{PA}$-formula

$$(\forall X)\Big(\big(X(0) \wedge (\forall x)(X(x) \Rightarrow X(Sx))\big) \implies X = (\forall x)X(x)\Big).$$

By adding it to PA, we would obtain the much stronger second-order theory PA_2.

Here X represents (any) unary relation; that is, a subset of the universe. The important distinction is that (3.3) provides an infinite collection of axioms, one for each subset of the universe that is *definable* by a $\mathcal{L}_{PA}$-formula φ.

This restriction is the source of PA's most profound properties and limitations. For example, PA_2 is categorical; that is, it has only one model (up to isomorphism) — the standard model $\mathcal{N}$. On the other hand, PA allows the existence of other non-standard models.

3.3.2 Models of Arithmetic

We have already mentioned that the *standard model* of PA is the $\mathcal{L}_{PA}$-structure $\mathcal{N} = (\omega, 0^{\mathcal{N}}, S^{\mathcal{N}}, +^{\mathcal{N}}, \cdot^{\mathcal{N}}, \leq^{\mathcal{N}})$, where the domain is the set ω, the interpretation of the symbol '0' is $0^{\mathcal{N}} = \varnothing$, the successor of x is $S^{\mathcal{N}}(x) = x \cup \{x\}$, and $+^{\mathcal{N}}$, $\cdot^{\mathcal{N}}$ and $\leq^{\mathcal{N}}$ are the operations of ordinal arithmetic restricted to ω.

Theorem 3.44. *There exist countable models of* PA *that are not isomorphic to* $\mathcal{N}$*.*

Proof sketch. By the Compactness Theorem. We extend $\mathcal{L}_{PA}$ with a new constant symbol c. Consider the theory $T = \mathsf{PA} \cup \{c > \bar{n} \mid n \in \omega\}$, where $\bar{n}$ is the $\mathcal{L}_{PA}$-term $S(S(\dots S(0)\dots))$ (n times). Any finite subset $T_0 \subset T$ is satisfiable: we take $\mathcal{N}$ as the model and interpret c as a standard natural number larger than any numeral $\bar{n}$ explicitly mentioned in T_0. By the Compactness Theorem, T has a model $\mathcal{M}$. This $\mathcal{M}$ must be a model of PA, but the interpretation of c is a "non-standard" number, an element larger than all standard elements $S^n(0)$. Thus, $\mathcal{M} \ncong \mathcal{N}$. $\square$

All countable non-standard models $\mathcal{M}$ share a common structure: they begin with an initial segment isomorphic to ω (the standard part), which is then followed by a collection of "blocks" of non-standard numbers. This "pathology" of PA is not merely set-theoretic, but also computational.

Theorem 3.45 (Tennenbaum, 1959)**.** *No countable non-standard model of* PA *is recursive.*

This implies that in any non-standard model $\mathcal{M}$, the operations $\oplus$ and $\otimes$ (the interpretations of $+$ and $\cdot$) are not computable functions. Even if the domain of $\mathcal{M}$ is ω, the operations themselves cannot be implemented by an algorithm. The induction schema, while syntactically "weaker" than its second-order counterpart, thus imposes enormous computational complexity on any structure that satisfies it, effectively isolating the standard model as the only computationally tractable one.

3.3.3 Gödel's Incompleteness Theorems

When proving the properties of a formal theory, the two most fundamental are consistency and completeness. A theory is *consistent* if it is free from contradictions, meaning it is impossible to prove both a statement ϕ and its negation $\neg\phi$ from its axioms; or equivalently, if it has a model. A consistent theory is *complete* if it has an "opinion" on every statement, meaning for every sentence ϕ in its language, the theory can prove either ϕ or $\neg\phi$. If it cannot do either, it is said to be *incomplete*, and ϕ is said to be *independent*. Equivalently, ϕ holds in some models of the theory but does not hold in others.

Probably the most influential result linking these two concepts together with PA are the famous Incompleteness Theorems, published by Kurt Gödel [6] [11] in 1931. Veritasium has an amazing video [23] that provides an intuitive explanation of this topic. We provide only a simplified explanation of these profound results; for more details and proofs, refer to [3].

Despite its limitations, PA is a remarkably powerful theory. Its expressive power is sufficient to define all computable (recursive) functions. This strength is the key to PA's own undoing. It allows for the *arithmetization of syntax* (Gödel numbering), whereby the syntax of $\mathcal{L}_{PA}$—terms, formulas, proofs—can be uniquely encoded as natural numbers (using prime numbers). Syntactic operations (like substitution) and relations (like "is a proof of") become recursive functions and relations on these numbers. Crucially, this allows for the creation of a provability predicate.

Definition 3.46 (Provability Predicate)**.** There exists an $\mathcal{L}_{PA}$-formula $\mathrm{Prov}_{\mathsf{PA}}(x)$ such that for any sentence ϕ it holds that $\mathsf{PA} \vdash \phi \Leftrightarrow \mathrm{Prov}_{\mathsf{PA}}(\ulcorner\phi\urcorner)$. Here, $\ulcorner\phi\urcorner$ denotes the Gödel number of ϕ. The formula $\mathrm{Prov}_{\mathsf{PA}}(\ulcorner\phi\urcorner)$ is: "there exists x such that x is the Gödel number of a proof of the sentence with Gödel number $\ulcorner\phi\urcorner$."

This predicate allows the theory to "talk about" its own provability, leading directly to sentences that self-reference and assert their own unprovability.

Theorem 3.47 (Gödel's First Incompleteness Theorem, 1931)**.** *If* PA *is consistent, then it is incomplete.*

Proof sketch. Consider a sentence $\mathfrak{g}$ (the *Gödel sentence*) saying: "there is no x such that x is the Gödel number of a proof of the sentence with Gödel number $\ulcorner\mathfrak{g}\urcorner$." Notice that $(\mathsf{PA} \vdash \mathfrak{g}) \iff \neg\mathrm{Prov}_{\mathsf{PA}}(\ulcorner\mathfrak{g}\urcorner)$. Hence if $\mathsf{PA} \vdash \mathfrak{g}$, then PA is inconsistent. Therefore, if PA is consistent, then $\mathsf{PA} \nvdash \mathfrak{g}$, and it is incomplete. □

As a corollary of this theorem, Gödel achieved his second result.

Theorem 3.48 (Gödel's Second Incompleteness Theorem, 1931)**.** PA *cannot prove its own consistency.*

[6] Gödel's (1906–1978) life is a fascinating story. Born in Brno, he left for Vienna at the age of eighteen to study mathematics and logic. At twenty-four, he proved his incompleteness theorems—work that formed the basis of his doctoral dissertation. He later emigrated to the United States following the rise of Nazism. Albert Einstein regarded Gödel as the greatest logician since Aristotle and once remarked that the only reason he went to his office was to have the privilege of walking home with Gödel. Yet Gödel's life was not without darkness: he struggled with psychological illness throughout adulthood and ultimately died of self-starvation, driven by the paranoid belief that he was being poisoned. Perhaps the most detailed account of Gödel's life (as of the writing of this text) can be found in [2].

Proof sketch. Let $\mathrm{Con}(\mathsf{PA})$ be the $\mathcal{L}_{PA}$-sentence $\neg\mathrm{Prov}_{\mathsf{PA}}(\ulcorner\bot\urcorner)$ (where $\bot$ is a contradiction, e.g., $0 = S0$). That is, $\mathrm{Con}(\mathsf{PA})$ is true if and only if PA is consistent. Because Gödel formalized the entire proof of the previous theorem in PA (using Gödel numbers), his first theorem can be expressed as

$$\mathsf{PA} \vdash \big(\mathrm{Con}(\mathsf{PA}) \implies \neg\mathrm{Prov}_{\mathsf{PA}}(\ulcorner\mathfrak{g}\urcorner)\big).$$

This together with the equivalence $(\mathsf{PA} \vdash \mathfrak{g}) \iff \neg\mathrm{Prov}_{\mathsf{PA}}(\ulcorner\mathfrak{g}\urcorner)$ gives

$$\mathsf{PA} \vdash \big(\mathrm{Con}(\mathsf{PA}) \implies \mathfrak{g}\big).$$

Now, suppose for contradiction that PA could prove its own consistency. Combining this with the last formula gives $\mathsf{PA} \vdash \mathfrak{g}$, but this is a contradiction, since (the end of the previous proof) if PA is consistent, then $\mathsf{PA} \nvdash \mathfrak{g}$. $\square$

Gödel's original formulation of these theorems did not, in fact, talk about PA, but about a system he called P, a close relative of PA. Gödel then had to make a philosophical assumption. He argued that any other system "related" to, and at least as strong as P (and therefore capable of arithmetic), would also be capable of producing a Gödel sentence $\mathfrak{g}$; thus, his incompleteness theorems would apply to this system as well. This was a strong, intuitive argument, but he could not formally prove it.

The missing piece was provided in 1936 by Alan Turing [22], who formalized the notion of computability using the Turing machine, which made a formal proof of Gödel's conjecture possible.

Theorem 3.49 (Generalized Gödel's Incompleteness Theorems)**.** *For any consistent, recursively axiomatized theory T, it holds that:*

(1) *If T is an extension of Robinson arithmetic* Q*, then T is incomplete.*

(2) *If T is an extension of Peano arithmetic* PA*, then T cannot prove its own consistency.*

Remark. Recursively axiomatized means that there is an algorithm that, for every input formula φ, halts and answers whether φ is an axiom of T. The condition that T is an extension of Q (or PA) essentially means that T is at least as powerful as Q (or PA). For example, PA is an extension of Q.

Corollary 3.50. *It is impossible to prove the consistency of* ZFC *inside* ZFC*.*

Remark. An example of an independent statement in ZFC is the famous continuum hypothesis (CH), claiming that there is no set x such that $\omega \prec x \prec \mathcal{P}(\omega)$. Similarly, AC can be shown to be independent in ZF, meaning that if ZF is consistent, then ZFC is as well. In 1940, Gödel showed that neither AC can be disproved from ZF, nor CH from ZFC, by constructing the *constructible universe*, a model of ZF in which both AC and CH hold. This model begins with the empty set and adds only those sets that are definable from previous ones, thus forming the minimal universe compatible with the axioms. Later, in 1963, Cohen showed that CH cannot be proven from ZFC by developing the method of *forcing*, which allowed him to construct a model of ZFC in which CH fails. Through a different forcing argument, he likewise obtained a model of ZF that violates AC.

3.3.4 Consistency and the Connection with ε_0

Gödel's second theorem seems to place us in a difficult position: a consistency proof for PA must employ principles that transcend PA itself. While ZFC is far stronger than PA and easily proves $\mathrm{Con}(\mathsf{PA})$ (by exhibiting the model $\mathcal{N}$), this isn't a very "unilluminating" result. PA is a "finitary" theory, while ZFC is a wildly "infinitary" theory (it assumes the existence of various vast infinities). By proving the consistency of PA in ZFC, we base our proof on the assumption that ZFC is consistent. It would be better to find a weaker system that is still capable of proving $\mathrm{Con}(\mathsf{PA})$.

Theorem 3.51 (Gentzen's Consistency Proof [7] , 1936)**.** *The consistency of* PA *is provable in Primitive Recursive Arithmetic* PRA *(which by itself is weaker than* PA*), augmented with a schema for transfinite induction up to the ordinal* ε_0*.*

Gentzen's proof precisely identified the "transcendent principle" required. Recall from Section 3.2.4 that ε_0 is the first fixed point of the ordinal function $\alpha \mapsto \omega^\alpha$. It is the limit of the sequence $\omega, \omega^\omega, \omega^{\omega^\omega}, \ldots$. Gentzen's result, $\mathsf{PRA} + \mathrm{TI}(\varepsilon_0) \vdash \mathrm{Con}(\mathsf{PA})$, thus establishes two facts:

(a) Proving the consistency of PA does not require the full power of ZFC; only transfinite induction up to a countable ordinal. That is, the assumption that ε_0 contains no infinite decreasing chains.

(b) The principle of transfinite induction up to ε_0, $\mathrm{TI}(\varepsilon_0)$, must be unprovable in PA (lest PA prove its own consistency).

Gentzen also showed that using any smaller ordinal $\alpha < \varepsilon_0$ is not enough. The idea of the proof is closely related to the fact that every ordinal $\alpha < \varepsilon_0$ has a (finite) hereditary Cantor normal form, while ε_0 does not. Hereditary CNF simply means that if any of the powers γ_i are ordinals larger than ω, then we express them in CNF as well. For example:

$$\alpha = \omega^{\omega^{\omega+1}+\omega^2\cdot 3+5} + \omega^{\omega\cdot 2+1} + \omega\cdot 2 + 7.$$

But ε_0 cannot be represented by a finite hereditary CNF, since $\varepsilon_0 = \omega^{\varepsilon_0}$ is its CNF, which is self-referential.

This calibrates the strength of PA with extraordinary precision. The collected strength of PA's infinite induction schema is exactly equivalent to the single principle of transfinite induction up to (but not including) ε_0. This is formalized in the concept of the *proof-theoretic ordinal.*

Theorem 3.52. *The* proof-theoretic ordinal *of* PA *is* $|\mathsf{PA}| = \varepsilon_0$*.*

This theorem has a twofold meaning that we can understand intuitively:

(a) **What PA *can* prove:** PA is strong enough to prove the well-foundedness of any *recursive well-ordering* $<_R$ on ω with order-type $\alpha < \varepsilon_0$.

(b) **What PA *cannot* prove:** PA is *not* strong enough to prove the well-foundedness of any recursive well-ordering $<_R$ on ω with order-type $\alpha \geq \varepsilon_0$.

[7] See [18] for a modern version of the proof. Moreover, [7] provides an alternative view on this result, and talks about the consistency of PA in general.

Here, a "recursive ordering" is simply a relation $<_R$ on ω that a computer can check. Well-foundedness is the arithmetical statement that this $<_R$ admits no infinite descending sequences $a_0 >_R a_1 >_R a_2 >_R \cdots$. Proving well-foundedness is thus equivalent to proving that transfinite induction "works" for that ordering.

A point of confusion here might be the fact that any well-ordering is well-founded. But PA does not know that $<_R$ is a well-ordering; it only receives an "object," $<_R$, together with the instructions: "prove that the ordering you received is well-founded."

Therefore, PA can formalize proofs by transfinite induction up to any ordinal $\alpha < \varepsilon_0$, but it cannot justify the principle of transfinite induction up to ε_0 itself.

3.3.5 Beyond Predicativity: ID_n and ATR_0

The discovery of set-theoretic paradoxes (such as Russell's and Burali-Forti's) in the early 20th century triggered a profound *Grundlagenkrise*, or foundational crisis, in mathematics. The naive assumption that any property $\phi(x)$ could define a set $\{x \mid \phi(x)\}$ was proven contradictory.

This led to several philosophical responses. While most of mathematics eventually adopted the axiomatic approach of ZFC (which restricted the formation of sets), a significant objection was raised by mathematicians like Poincaré and, most systematically, Hermann Weyl. They argued that the core problem was the use of *impredicative definitions*—definitions that define an object S by quantifying over a totality T that already includes S. Weyl, in his 1918 monograph *Das Kontinuum* (for english translation see [26]), argued that such definitions were circular, and he sought to rebuild analysis on a "safe" *predicative* basis, using only definitions that build "from the ground up" without reference to a completed, all-encompassing totality.

Gentzen's work establishes ε_0 as the limit of PA, a theory that is fundamentally predicative. Its induction schema applies to properties definable using only quantifiers over the natural numbers, a totality which is considered "fixed" or "already given." To obtain theories stronger than PA, one must introduce principles that are, in some sense, impredicative—principles that define new objects by quantifying over a totality that includes the object being defined. For example, the supremum of a set is defined as its least upper bound. In this case, the totality is the set of all upper bounds, and since the supremum is itself an upper bound, it is a member of that totality.

The following is a gross simplification of the results that stem from this. Refer to [20] for formal proofs and precise formulations. Additionally, [8] explores the rich history of this field and provides a more high-level perspective.

Definition 3.53 (Inductive Definitions, ID_1)**.** An *arithmetical inductive definition* is given by an arithmetical formula $\mathcal{O}(X, x)$, where X is a set variable. This formula is *monotone* if $\forall x(\mathcal{O}(X, x) \to \mathcal{O}(Y, x))$ whenever $X \subseteq Y$. Such a formula $\mathcal{O}$ defines a set $I_{\mathcal{O}}$ as its *least fixed point*:

$$I_{\mathcal{O}} := \bigcap \{X \subseteq \omega \mid \forall x(\mathcal{O}(X, x) \to x \in X)\}$$

This definition is impredicative, as $I_{\mathcal{O}}$ is defined by quantifying over a collection (all subsets X satisfying the closure condition) which itself contains $I_{\mathcal{O}}$. The theory ID_1 (First-Order Arithmetic with one Inductive Definition) is PA augmented with an axiom schema asserting that for any such arithmetical formula $\mathcal{O}$, the least fixed point $I_{\mathcal{O}}$ exists, and (crucially) allowing arithmetical induction to be applied to formulas containing $I_{\mathcal{O}}$.

This addition represents an enormous leap in proof-theoretic strength, measured precisely by the next level of the Veblen hierarchy, which we encountered in Section 3.2.4.

Theorem 3.54 (Proof-Theoretic Ordinal of ID_1)**.** *The proof-theoretic ordinal of* ID_1 *is the first fixed point of the ε-function (the first zeta-number):*

$$|\mathsf{ID}_1| = \zeta_0 = \varphi_2(0) = \sup\{\epsilon_0, \epsilon_{\epsilon_0}, \epsilon_{\epsilon_{\epsilon_0}}, \dots\}$$

Remark (The ID_n Hierarchy). This process can be iterated. The theory ID_2 allows for inductive definitions that arithmetically refer to the set I_1 (defined by the inductive definiton in ID_1), ID_3 iterates this again, and so on. This hierarchy of theories is measured precisely by the Veblen hierarchy:

$$|\mathsf{ID}_n| = \varphi_{n+1}(0)$$

The theory $\mathsf{ID}_{<\omega} = \bigcup_{n\in\omega} \mathsf{ID}_n$ is the theory that allows for any finite iteration of this inductive definition process. Its proof-theoretic strength is the supremum of all these ordinals.

When Poincaré and Weyl set out to rebuild mathematics without impredicativeness, the big question became: "How much mathematics can we *actually* recover using only predicative methods?" In the 1960s, Solomon Feferman and Kurt Schütte (independently) set out to find the precise answer. They formalized what "predicative mathematics" means and created a system often called ATR_0, or Arithmetical Transfinite Recursion.

Definition 3.55 (Arithmetical Transfinite Recursion, ATR_0)**.** The theory ATR_0 is a subsystem of second-order arithmetic. Its central axiom asserts that for any arithmetical formula $\mathcal{O}(x, X)$ and any countable well-ordering $\prec$, one can define a new set H by transfinite recursion along $\prec$. That is, one can iterate an arithmetical definition "along" any countable well-ordering. This theory is considered to be the formal upper bound of predicative reasoning.

Feferman (in 1964, [10]) and Schütte (in 1965, [19]) then went on to prove that these two distinct approaches—iterating inductive definitions a finite number of times, and formalizing predicative mathematics—are, in fact, equivalent in strength.

Theorem 3.56 (Feferman–Schütte, 1964–1965)**.** *The theories* $\mathsf{ID}_{<\omega}$ *and* ATR_0 *are proof-theoretically equivalent. Their shared proof-theoretic ordinal is the* Feferman–Schütte ordinal*,* Γ_0*.*

$$|\mathsf{ID}_{<\omega}| = |\mathsf{ATR}_0| = \Gamma_0 = \sup_{n<\omega}\{\varphi_n(0)\}$$

The calibration of these impredicative theories was a monumental achievement of post-Gentzen proof theory. The independent work of Solomon Feferman and Kurt Schütte in the 1960s established Γ_0 as the precise ordinal limit of predicative analysis. This shows that Γ_0 is the first ordinal that cannot be proven to be well-founded by predicative means, just as ε_0 is the first ordinal that cannot be proven well-founded by the finitistic (in Gentzen's sense) means of PA.

3.4 Goodstein Sequences and the Hydra Game

Gödel's independent sentences, $\mathfrak{g}$ and Con(PA), are meta-mathematical statements, not "natural" theorems of number theory. For decades, it was an open question whether any "ordinary" theorem of arithmetic or combinatorics was unprovable in PA, leading to speculation that Gödel's Incompleteness Theorem would not have meaningful implications to practical mathematics.

However, in 1977, Parris and Harrington [17] showed that a very natural variation of Ramsey's Theorem was true, but not provable in PA. Five years later, in 1982, Kirby and Paris [16] showed that Goodstein theorem, a statement purely about sequences of natural numbers, cannot be proven in PA either.

Goodstein proved his theorem in 1944 [12] by mapping each term in a Goodstein sequence to a transfinite ordinal less than ε_0. He then showed that this corresponding sequence of ordinals is strictly decreasing. The theorem's claim that every Goodstein sequence eventually terminates is thus a direct consequence of the well-foundedness of ε_0 — since there are no infinite decreasing chains, the sequence of ordinals has to terminate, and the corresponding Goodstein sequence with it.

Goodstein's proof utilized transfinite methods, and for a long time, it was not clear whether a finitary proof—one formalizable in PA—might also exist. Kirby and Paris proved that no such finitary proof is possible. They showed that Goodstein's Theorem, while true in $\mathcal{N}$ (and provable in ZFC), is unprovable in PA. They did so by proving that

$$\mathsf{PA} \vdash \big(\mathrm{GT} \implies \mathrm{Con}(\mathsf{PA})\big).$$

Combining this with Gödel's second incompleteness theorem, $\mathsf{PA} \nvdash \mathrm{Con}(\mathsf{PA})$, they arrived at the conclusion that $\mathsf{PA} \nvdash \mathrm{GT}$.

This is an astonishing result. As shown in [21], PA is equivalent to the theory of finite sets; that is, ZFC with the axiom of infinity replaced by the axiom "there are no limit ordinals." From this, one can prove that all sets are finite. The Kirby–Paris theorem asserts that we need to reach out to infinity to prove a statement about finite sets.

The second theorem presented in Kirby and Paris's 1982 paper establishes an analogous result, this time showing that a statement about the Hydra game is unprovable in PA.

Goodstein Sequences

The Hydra Game

Sources

This document is mostly my notes from the class NMAI074 taught at MFF CUNI by doc. Kynčl. The web of the course is HERE. The course mostly follows parts of the second and third chapters of [1], which is in Czech. My notes from the introductory set theory course can be found HERE, also in Czech.

If you found any mistakes or errors, please contact me at smolikj@matfyz.cz.

[1] Petr Balcar Bohuslav a Štěpánek. *Teorie množin.* Vydání 2., opravené a rozšířené. Praha: Academia, 2001. ISBN: 80-200-0470-X.

[2] Stephen Budiansky. *Journey to the Edge of Reason: The Life of Kurt Gödel.* W. W. Norton, 2021. ISBN: 9781324005452.

[3] Jakub Bulín. *NAIL062 Propositional and Predicate Logic: Lecture Notes.* 2025, pp. 149–164. URL: `https://github.com/jbulin-mff-uk/nail062/raw/main/lecture/lecture-notes/lecture-notes.pdf` (visited on 11/06/2025).

[4] Jochen Burghardt. *Matchstick representation of ordinal numbers up to ω^ω.* 2023. URL: `https://commons.wikimedia.org/wiki/File:Omega-exp-omega-normal.pdf` (visited on 11/04/2025).

[5] Jochen Burghardt. *Order Type Examples.* 2019. URL: `https://commons.wikimedia.org/wiki/File:OrderTypeExamples.pdf` (visited on 11/04/2025).

[6] Pop-up casket. *A spiral representation of ordinals less than ω^ω.* 2012. URL: `https://en.wikipedia.org/wiki/File:Omega-exp-omega-labeled.svg` (visited on 11/04/2025).

[7] Timothy Y Chow. "The consistency of arithmetic". In: *The Mathematical Intelligencer* 41.1 (2019), pp. 22–30. URL: `https://arxiv.org/pdf/1807.05641`.

[8] Walter Dean and Sean Walsh. "The prehistory of the subsystems of second-order arithmetic". In: *The Review of Symbolic Logic* 10.2 (2017), pp. 357–396. URL: `https://arxiv.org/pdf/1612.06219`.

[9] Herbert B Enderton. *Elements of set theory.* Gulf Professional Publishing, 1977, pp. 195–196.

[10] Solomon Feferman. "Systems of predicative analysis1". In: *The Journal of Symbolic Logic* 29.1 (1964), pp. 1–30.

[11] Kurt Gödel. "On Formally Undecidable Propositions of Principia Mathematica and Related Systems I". In: *Kurt Gödel: Collected Works: Volume I Publications 1929-1936.* OUP Oxford, 1986, pp. 144–195. ISBN: 0-19-503964-5.

[12] Reuben Louis Goodstein. "On the restricted ordinal theorem". In: *The Journal of Symbolic Logic* 9.2 (1944), pp. 33–41. DOI: `10.2307/2268019`.

[13] Gro-Tsen and IkamusumeFan. *A graphical "matchstick" representation of the ordinal ω^2.* 2015. URL: `https://commons.wikimedia.org/wiki/File:Ordinal_ww.svg` (visited on 11/04/2025).

[14] Joel David Hamkins. *Transfinite recursion as a fundamental principle in set theory*. Accessed: 2025-11-08. 2014. URL: https://jdh.hamkins.org/transfinite-recursion-as-a-fundamental-principle-in-set-theory/.

[15] Karel Hrbáček and Tomáš Jech. *Introduction to set theory*. eng. Third edition, revised and expanded. Pure and applied mathematics. A series of monographs and textbooks ; 220. Boca Raton: Taylor & Francis, 1999. ISBN: 0-8247-7915-0.

[16] Laurie Kirby and Jeff Paris. "Accessible independence results for Peano arithmetic". In: *Bulletin of the London Mathematical Society* 14.4 (1982), pp. 285–293. DOI: 10.1112/blms/14.4.285.

[17] Jeff Paris and Leo Harrington. "A mathematical incompleteness in Peano arithmetic". In: *Handbook of Mathematical Logic, edited by J. Barwise*. North-Holland, 1977, pp. 1133–1142.

[18] Michael Rathjen. "The art of ordinal analysis". In: *Proceedings of the International Congress of Mathematicians*. Vol. 2. European Mathematical Society. 2006, pp. 45–69.

[19] Kurt Schütte. "Predicative well-orderings". In: *Studies in Logic and the Foundations of Mathematics*. Vol. 40. Elsevier, 1965, pp. 280–303.

[20] Stephen George Simpson. *Subsystems of second order arithmetic*. Vol. 1. Cambridge University Press, 2009.

[21] Will Sladek. *The termite and the tower: Goodstein sequences and provability in PA*. 2007. URL: https://andrescaicedo.wordpress.com/wp-content/uploads/2017/09/sladekgoodstein.pdf.

[22] Alan Mathison Turing et al. "On computable numbers, with an application to the Entscheidungsproblem". In: *J. of Math* 58.345-363 (1936), p. 5.

[23] Veritasium. *Math's Fundamental Flaw*. 2021. URL: https://www.youtube.com/watch?v=HeQX2HjkcNo (visited on 11/06/2025).

[24] Veritasium. *The Man Who Almost Broke Math (And Himself...) - Axiom of Choice*. 2025. URL: https://www.youtube.com/watch?v=_cr46G2K5Fo (visited on 11/08/2025).

[25] Vsauce. *How To Count Past Infinity*. 2016. URL: https://www.youtube.com/watch?v=SrU9YDoXE88 (visited on 11/04/2025).

[26] Hermann Weyl. *The Continuum: a critical examination of the foundation of analysis ; translated by Stephen Pollard & Thomas Bole*. Courier Corporation, 1994.

[27] Wikipedia. *Hausdorff maximal principle — Wikipedia, The Free Encyclopedia*. 2025. URL: https://en.wikipedia.org/w/index.php?title=Hausdorff%5C%20maximal%5C%20principle&oldid=1300391387 (visited on 10/07/2025).